

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 1, 2005, 17:24:23 ; Search time 4399 Seconds
(without alignments)
2478.387 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSKVKVKEFMKVRMEG.....EDYIVQEYVTEGRHHLFL 225

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Deiop 6.0 , Deilext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/US10006922/runat_30062005.110616.19088/app.query.fasta_1.391
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10006922.cgn 1 1 3731 @runat_30062005.110616.19088 -NCPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1214	100.0	678	6 AR183915 Sequence
2	1214	100.0	678	6 AX172854 Sequence
3	1214	100.0	678	6 AX207715 Sequence
4	1214	100.0	678	6 AX233581 Sequence

5	1214	100.0	678	6 AX370404 Sequence
6	1214	100.0	678	6 AX824725 Sequence
7	1214	100.0	859	3 AF168419 Discosoma
8	1214	100.0	859	6 AX463698 Sequence
9	1210	99.7	681	6 AX233584 Sequence
10	1210	99.7	713	6 AX233627 Sequence
11	1210	99.7	723	6 AR527331 Sequence
12	1210	99.7	2721	6 CQ882115 Sequence
13	1210	99.7	2772	6 CQ882117 Sequence
14	1210	99.7	4692	6 AX463702 Sequence
15	1210	99.7	8893	6 AX823860 Sequence
16	1210	99.7	8811	12 AY569779 Cloning v
17	1210	99.7	9320	6 AX663075 Sequence
18	1210	99.7	12404	12 AY569780 Cloning v
19	1207	99.4	678	6 AX370406 Sequence
20	1196	98.5	666	6 AX348043 Sequence
21	1196	98.5	666	6 AX353910 Sequence
22	1196	98.5	711	3 AY679107 Discosoma
23	1193	98.3	678	3 AF545828 Discosoma
24	1191	98.1	678	6 AX370408 Sequence
25	1191	98.1	678	6 AX824732 Sequence
26	1191	98.1	921	3 AY679106 Discosoma
27	1186	97.7	675	6 AX824731 Sequence
28	1186	97.7	1050	6 AX666133 Sequence
29	1186	97.7	4488	6 CQ849509 Sequence
30	1186	97.7	6423	12 AY613997 Cloning v
31	1186	97.7	7616	6 CQ849511 Sequence
32	1184	97.5	898	6 AX686888 Sequence
33	1165	96.0	10141	12 AY342347 Red H-Pel
34	1165	96.0	10276	12 AY342348 Red H-Sti
35	1165	96.0	10481	12 AY490568 UAS-Red S
36	1121	92.3	681	12 AFS06025 Synthetic
37	1121	92.3	1395	12 AFS06026 Synthetic
38	1119	92.2	678	6 AX824729 Sequence
39	1085.5	89.4	876	3 AF272711 Discosoma
40	1085.5	89.4	876	3 AX686894 Sequence
41	1021	84.1	678	12 AFS06027 Synthetic
42	1021	84.1	696	12 ABI66761 Synthetic
43	805	66.3	669	6 AX699925 Sequence
44	804	66.2	666	3 AY646074 Actopora
45	800	65.9	663	6 AX699859 Sequence

ALIGNMENTS

RESULT 1	AR183915	Sequence 6 from patent US 6342379.	678 bp	DNA	linear	PAT 20-APR-2002
LOCUS	AR183915	Sequence 6 from patent US 6342379.				
DEFINITION	AR183915	Sequence 6 from patent US 6342379.				
ACCESSION	AR183915	Sequence 6 from patent US 6342379.				
VERSION	AR183915.1	GI:20227884				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 678)					
AUTHORS	Tsien,R.Y. and Gonzalez,J.E. III.					
TITLE	Detection of transmembrane potentials by optical methods					
JOURNAL	Patent: US 6342379-A 6 29-JAN-2002;					
FEATURES	Location/Qualifiers					
source	1..678					
organism	/organism="unknown"					
mol_type	/mol_type="unassigned DNA"					

Alignment Scores:	2.62e-119	Length:	678
Pred. No.:	1214.00	Matches:	225
Score:	100.00%	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	100.00%	Gaps:	0
DB:	6		

Handwritten notes: "Xgapop" and "20 Apr 2002" with a signature.

US-10-006-922A-12 (1-225) x AR183915 (1-678)

Qy	1	MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Db	1	ATGAGGTCTTCCAAGAAATGTTATCAAGAGATTCTAGAGGTTTAAGGTTTCGCATGGAAGGA	60
Qy	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Db	61	ACGGTCNATGGGCACGAGTTTGGNAATAGAGCCGAGGAGNGGGAGCCATACGAAAGC	120
Qy	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaIleAspIle	60
Db	121	CACAATACCGTAAAGCTTAAAGTAACCAAGGGGGGACCTTTGCCATTTCGTTGGGATATT	180
Qy	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Db	181	TTGTGCACCAATTTTCAGTATGGGAAGAGGTATATGTCAAGACACCTGCCGACATACCA	240
Qy	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Db	241	GACTATAAAAAGCTGTCTATTTCTCAAGGATTTTAAATGGGAAGGGTCATGAACCTTTGAA	300
Qy	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Db	301	GACGGTGGCGTCGTACTGTGTAACCCAGAGATTCACAGTTTGACAGGATGGCTGTTTCATCTAC	360
Qy	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Db	361	AAGGTCNAAGTTCATTGGCGTGAACCTTTCTTCGATGGACCTGTTATGCCAAGAAGACAC	420
Qy	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Db	421	ATGGGCTGGGAAGCCAGCACCTGAGCGTTTGTATCTCTGATGGCGTGTGTTGAAAGAGAG	480
Qy	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
Db	481	ATTCTAAGGCTCTGNAGCTGAAAGCCGGTGGTCATTACCTAGTTTGAATTTCAAAAGTATT	540
Qy	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Db	541	TACATGGCAAGAAAGCCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT	600
Qy	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Db	601	ATNACAAAGCCCAACGAAGACTATACAAATCGTTGAGCAGCTATGAAGACCCGAGGACGC	660
Qy	221	HisHisLeuPheLeu	225
Db	661	CACCATCTGTCTCT	675

1994

Alignment Scores:

Alignment Scores:			
Pred. No.:	2,628-119	Length:	678
Score:	1214.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-006-922A-12 (1-225) x AX172854 (1-678)

Qy	1	MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Db	1	ATGAGGTCTTCCCAAGAAATGTTATCAAGGAGCTTCATGAGGTTTAAAGTTTCGATCGGAAGGA	60
Qy	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly	40
Db	61	ACGGTCANTGGGCACGAGTTTGAAATAGAAAGGCGAAGGAGGGGAGGCCCATACGAAGGC	120
Qy	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Db	121	CACAATACCGTAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT	180
Qy	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Db	181	TTGTGCACCACAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCCCTGCGGCACATACCA	240
Qy	81	AspTyrIleLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Db	241	GACTATAAAGAAGCTGTCATTTCTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA	300
Qy	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Db	301	GACGGTGGCGTCGTACTGTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTAC	360
Qy	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Db	361	AAGGTCAAGTTCATTTGGCGTGAACCTTTCTCTCCGATGGACCTGTATATGCAAAAAGAGACA	420
Qy	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Db	421	ATGGCGTGGGAAGCCAGCACTGAGCGTTTGATCTCTCGTATCGCGTGTGTAAGGAGAG	480
Qy	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
Db	481	ATTCATAAGGCTCTGAAGCTGAAAGCGTGGTGCATTACCTAGTTGAATTCAAAAGTATT	540
Qy	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Db	541	TACATGGCAAGAAGCCGTGTGCAGCTACCAAGGGTACTACTATGTGACTCCAACTGGAT	600
Qy	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Db	601	ATACCAAGCCACCAACGAAGACTATACANTCGTTGAGCAGTATGAAGAACCAGGAGCCG	660
Qy	221	HisHisLeuPheLeu	225
Db	661	CACCATCTGTTCTCT	675
RESULT 3			
AX207715			
LOCUS		678 bp	linear
DEFINITION	Sequence 13 from Patent WO0157242.		PAT 31-AUG-2001
ACCESSION	AX207715		
VERSION	AX207715.1	GI:15422399	
KEYWORDS			
SOURCE			
ORGANISM			
	Discosoma sp.		
	Discosoma sp.		
	Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;		
	Corallimorpharia; Discosomatidae; Discosoma.		
REFERENCE			
AUTHORS	Stack, J.H., Whitney, M., Cubitt, A.B. and Pollok, B.A.		
TITLE	Methods of protein destabilization and uses thereof		
JOURNAL	Patent: WO 0157242-A 13 09-AUG-2001;		

ORIGIN

```
FEATURES          Aurora Biosciences Corporation (US)
SOURCE            Location/Qualifiers
                  1. .678
                  /organism="Discosoma sp."
                  /mol_type="unassigned DNA"
                  /db_xref="taxon:86600"

ORIGIN
Alignment Scores:      2.62e-119      Length:      678
Pred. No.:            1214.00          Matches:    225
Score:                100.00%          Conservative: 0
Percent Similarity:   100.00%          Mismatches:  0
Best Local Similarity: 100.00%          Indels:      0
Query Match:         100.00%          Gaps:        6
DB:
US-10-006-922A-12 (1-225) x AX207715 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGCTCTTCCAGAGATGTTATCAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGA 60
QY 21 ThrValAsnGlyHisGluPheGluGlyGluGlyGluGlyGluGlyGluGlyGluGly 40
DB 61 ACGGTCATGCGCAGCAGTTTGAATACAGCGCGAAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyGlyGlyGlyGlyGlyGlyGly 60
DB 121 CACATACCGTAAAGCTTAAGTAAAGCAAGAGTATATGTCAGACACCCCTGCCGACATACCA 240
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValThrLysValThrLysHisProAlaAspIlePro 80
DB 181 TTGTCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGACACCCCTGCCGACATACCA 240
QY 81 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
DB 301 GACGCTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
DB 361 AAGGTCAGATTCATTGGCGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
DB 421 ATGGCTGGGAAGCAGCAGCTGAGCGTTTGTATCTCTGATGGCGTGTGGAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
DB 481 ATTCATAGGCTCTGAAGCTGAAAGACGTCGTCATTACCTAGTTGAATTCAAAGATATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
DB 541 TACATGGCAAGAGACCTGTCAGCTACCGGCTACTATGTTGATCTCCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
DB 601 ATACAGCCACACAGACATATACATCTGTTGAGCAGTATGAAAGAACCGGAGGAGCG 660
QY 221 HisHisLeuPheLeu 225
DB 661 CACCATCTGTTCCCTT 675

RESULT 4
AX233581
LOCUS          AX233581          678 bp          DNA          linear          PAT 11-SEP-2001
DEFINITION    Sequence 5 from Patent WO0162919.
ACCESSION     AX233581
VERSION       AX233581.1
KEYWORDS      GI:15593305
```

```
SOURCE            Discosoma sp.
ORGANISM         Discosoma sp.
REFERENCE        1
AUTHORS          Nelson, D., Zamaira, E. and Tsien, R.
TITLE            Modified fluorescent proteins
JOURNAL          Patent: WO 0162919-A 5 30-AUG-2001;
                  Aurora Biosciences Corporation (US)
FEATURES          Location/Qualifiers
SOURCE            1. .678
                  /organism="Discosoma sp."
                  /mol_type="unassigned DNA"
                  /db_xref="taxon:86600"
                  /note="red"

ORIGIN
Alignment Scores:      2.62e-119      Length:      678
Pred. No.:            1214.00          Matches:    225
Score:                100.00%          Conservative: 0
Percent Similarity:   100.00%          Mismatches:  0
Best Local Similarity: 100.00%          Indels:      0
Query Match:         100.00%          Gaps:        6
DB:
US-10-006-922A-12 (1-225) x AX233581 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGCTCTTCCAGAGATGTTATCAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGA 60
QY 21 ThrValAsnGlyHisGluPheGluGlyGluGlyGluGlyGluGlyGluGlyGluGly 40
DB 61 ACGGTCATGCGCAGCAGTTTGAATACAGCGCGAAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyGlyGlyGlyGlyGlyGlyGly 60
DB 121 CACATACCGTAAAGCTTAAGTAAAGCAAGAGTATATGTCAGACACCCCTGCCGACATACCA 240
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValThrLysValThrLysHisProAlaAspIlePro 80
DB 181 TTGTCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGACACCCCTGCCGACATACCA 240
QY 81 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
DB 241 GACTATATAAAGCTGTCATTTCTGAGGATTTAAATGGGAAGGTCATGAACCTTTGAA 300
QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
DB 301 GACGCTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
DB 361 AAGGTCAGATTCATTGGCGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
DB 421 ATGGCTGGGAAGCAGCAGCTGAGCGTTTGTATCTCTGATGGCGTGTGGAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
DB 481 ATTCATAGGCTCTGAAGCTGAAAGACGTCGTCATTACCTAGTTGAATTCAAAGATATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
DB 541 TACATGGCAAGAGACCTGTCAGCTACCGGCTACTATGTTGATCTCCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
DB 601 ATACAGCCACACAGACATATACATCTGTTGAGCAGTATGAAAGAACCGGAGGAGCG 660
QY 221 HisHisLeuPheLeu 225
DB 661 CACCATCTGTTCCCTT 675
```



```

Qy 161 IleHisIysAlaLeuLysLeuLysAspGlyHisTyrIleValGluPheLysSerIle 180
Db 481 ATCCACAAGGCCCTGAAGAGCGGGCCACCTACCTGGTGGAGTTCAAGTCCATC 540
Qy 181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAGCCGTCAGCTGCCCGCTACTACTACGTGGACTCCAACTGGAC 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACTCCCAACAGGAGCTACACCATCGTGAGCAGTACGAGCGCACCGAGGGCGC 660
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACCTGTTCTCTG 675

RESULT 7
AF168419
LOCUS AF168419 859 bp mRNA linear INV 27-JUL-2001
DEFINITION Discosoma sp. fluorescent protein FP583 mRNA, complete cds.
ACCESSION AF168419
VERSION AF168419.2 GI:7105733
KEYWORDS
SOURCE Discosoma sp.
ORGANISM Discosoma sp.
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE
1 (bases 1 to 859)
Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
Markelov,M.L. and Lukyanov,S.A.
Fluorescent proteins from nonbioluminescent Anthozoa species
Nat. Biotechnol. 17 (10), 969-973 (1999)
99436614
PUBMED 10504696
2 (bases 1 to 859)
Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
Markelov,M.L. and Lukyanov,S.A.
Direct Submission
Submitted (14-JUL-1999) Institute of Bioorganic Chemistry RAS,
Miklukho-Maklavya 16/10, Moscow 117871, Russia
3 (bases 1 to 859)
Matz,M.V., Fradkov,A.F., Labas,Y.A., Savitsky,A.P., Zaraisky,A.G.,
Markelov,M.L. and Lukyanov,S.A.
Direct Submission
Submitted (25-FEB-2000) Institute of Bioorganic Chemistry RAS,
Miklukho-Maklavya 16/10, Moscow 117871, Russia
REMARK
COMMENT On Feb 25, 2000 this sequence version replaced gi:6090866.
FEATURES
Location/Qualifiers
1..859
/organism="Discosoma sp."
/mol_type="mRNA"
/db_xref="taxon:86600"
. 54..731
/note="GFP-like protein; orange-red"
/codon_start=1
/product="fluorescent protein FP583"
/protein_id="AAF03369.1"
/db_xref="GI:6090867"
/translation="MRSKKNVKEFMFKVMEKTVNGHFEIEGEGEGRPYEGHNTV
KLKVTGGPFPFADWILSPQFGSKYVVKHPADIPDYKKLSPGPKWERVMNFEDG
GVWTVTDSLSLQDGCIFYKVKFIGNFPDGPVNMKTMGWEASTERLYPRDGLKGE
IHKALKDGGHGLVLEFKSIYMAKKVQLPGYVYVDSKLDITSHNEDYTYVEQYERTE
GRHLFL"

ORIGIN
Alignment Scores:
Pred. No.: 3.53e-119 Length: 859
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0

```

```

DB: 3 Gaps: 0
US-10-006-922A-12 (1-225) x AF168419 (1-859)
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 54 ATGAGGTCTTCCAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTTCGATGGAAGGA 113
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 114 ACGGTCAATGGGCAGGTTTGAATAGAAAGGCGCAAGAGAGAGGAGGAGCCATACGAAGGC 173
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTyrAspIle 60
Db 174 CACAATACCGTAAAGCTTAAGGTAAACAAGGGGGGACCTTTGCCATTTCGTTGGGATATT 233
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 234 TTGTCAACCAATTTTCAATATGGAAGCAAGGTATATGTCAAGCACCTTCGCGACATACCA 293
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnPheGlu 100
Db 294 GACTATAAAAAGCTGTCATTTCCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 353
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 354 GACGGTGGCGTCGTTACTGTAAACCCAGGATTCAGATTTCAGGATGGCTGCTTTTCATCTAC 413
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 414 AAGTCAAGTTCAATTCGGGTGAACCTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 473
Qy 141 MetGlyTyrProGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 474 ATGGGTCTGGGAAGCCAGCAGCTGAGCGTTTGTATCTCGTCATGCGCGTTTGAAGGAGAG 533
Qy 161 IleHisIysAlaLeuLysLeuLysAspGlyGlyHisTyrIleValGluPheLysSerIle 180
Db 534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTTCATTACCTAGTTGAATTCAAAGATT 593
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 594 TACATGCCAAGAAAGCTGTGACAGCTACCAAGGTACTACTATGTTGACTCCAAACTGGAT 653
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 654 ATAAACAGCCACAAACCAAGACTATACAATCGTTGAGCAGTATGAAGAAGACCGAGGAGCG 713
Qy 221 HisHisLeuPheLeu 225
Db 714 CACCATCTGTTCTCTT 728

RESULT 8
AX463698
LOCUS AX463698 859 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 12 from Patent WO0248338.
ACCESSION AX463698
VERSION AX463698.1 GI:21886457
KEYWORDS
SOURCE Discosoma sp.
ORGANISM Discosoma sp.
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE
1
Lichtenberg-Frat,H.
Yeast strain for testing the geno- and cytotoxicity of complex
environmental contamination
Patent: WO 0248338-A 12 20-JUN-2002;
Lichtenberg-Frat, Hella (DE)
Location/Qualifiers
1..859
/organism="Discosoma sp."
/mol_type="unassigned DNA"

```

```
LOCUS AX233584 681 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 8 from Patent WO0162919.
ACCESSION AX233584
VERSION AX233584.1 GI:15593307
KEYWORDS synthetic construct
ORGANISM synthetic construct
REFERENCE 1 other sequences; artificial sequences.
AUTHORS Nelson, D., Zamaira, E. and Tsien, R.
TITLE Modified fluorescent proteins
JOURNAL Patent: WO 0162919-A 8 30-AUG-2001;
Aurora Biosciences Corporation (US)
FEATURES
    source
    Location/Qualifiers
        1..681
            /organism="synthetic construct"
            /mol_type="unassigned DNA"
            /db_xref="taxon:32630"
            /note="Mutant Anthozoan red fluorescent protein"
        1..681
            /note="unnamed protein product"
            /codon_start=1
            /transl_table=11
            /protein_id="CAC69734.1"
            /db_xref="GI:15593308"
            /translation="MVRSSKNVTKBFMRKVRMEGTVNGHEPEIEGEGEGRPYEGHNT
            VKLVTKGGPLPFWADILSPQYQSGKSVYVGHADIPDYKKLSPFGKWRVVMFEDG
            GVVTVQDSSLDQCFIYKVFIGNFSPDGPVMQKTMGWEASTERLYPRDGVLKGE
            EHKALKDKGGHYLVEFKSIYNAKKPVQLPGYIYVDSKLDITSHNEDYTIQEYERT
            EGRHFL"
ORIGIN
Alignment Scores:
Pred. No.: 6,99e-119 Length: 681
Score: 1210.00 Matches: 224
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 99.67% Indels: 0
DB: 6 Gaps: 0
US-10-006-922A-12 (1-225) x AX233584 (1-681)
Qy 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 4 GTGAGGAGCAGCAAGAACGGTATCAAGAGTTTCATGAGGTTCAAGGTGCCCATGGAGGC 63
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 64 ACCGTGAACCGCCACGAGTTCGAGATCGAGGCGGAGGCGGAGGCGCTACGAGGC 123
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 124 CACAACACCGTGAAGCTTAAGGTGACCAAGGCGGCGGCGGCTTCGCTTCGGCATC 183
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 184 CTGAGCCCCCAGTTCAGTACGCGCAGCAAGGTGACGTGAAGCACCCTCCGCGCATCCCC 243
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 244 GACTACAAGAGCTGAGCTTCCCGGAGGCTTCAAGTGGGAGGAGGTGATGAACCTTCGAG 303
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 304 GACGGCGCGTGGTGACCGTGACCCAGGACAGCAGCCTCGAGGACGCGCTCTTCATCTAC 363
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 364 AAGTGAAGTTTCATCGCGGTGAACCTCCCGCAGCAGCGCCCGTGTGATGCAGAGAACACC 423
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 424 ATGGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGGAGCGGCTGTGAAAGGCGGAG 483
/db_xref="taxon:86600"
54_731
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD42147.1"
/db_xref="GI:21886458"
/translation="MRSSKNVTKBFMRKVRMEGTVNGHEPEIEGEGEGRPYEGHNTV
KLVTGGPLPFWADILSPQYQSGKSVYVGHADIPDYKKLSPFGKWRVVMFEDG
GVVTVQDSSLDQCFIYKVFIGNFSPDGPVMQKTMGWEASTERLYPRDGVLKGE
EHKALKDKGGHYLVEFKSIYNAKKPVQLPGYIYVDSKLDITSHNEDYTIQEYERT
GRHFL"
ORIGIN
Alignment Scores:
Pred. No.: 3,53e-119 Length: 859
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-006-922A-12 (1-225) x AX463698 (1-859)
Qy 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 54 ATGAGGTTCTTCCAAAGATGTTATCAAGAGTTTCATGAGTTTAAGTTTCGCATGAAGGA 113
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 114 ACGGTCAATGGGCACGAGTTTGAATATGAAGCGCAAGGAGGAGGAGGCCATACGAAGGC 173
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 174 CACAATACCGTAAAGCTTAAGGTAAACCAAGGCGGAGCCCTTGGCCATTTGCTTGGATATT 233
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 234 TTGTCAACCAATTTCAATATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 293
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 294 GACTATAAAGAGCTGTCTATTTCTGAAGGATTTAAATGGGAAGGCTCATGAACCTTTGAA 353
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 354 GACGTGGCGTCTTACTGTAAACCCAGGATTCAGATTTCAGATTGCGAGGATGGCTGTTTCATCTAC 413
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 414 AAGGTCAAGTTTCATTTGGCGTGAACTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 473
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 474 ATGGGCTGGGAAGCCAGCACCTGAGCGTTTGTATCTCTGATGGCGTGTGAAAGGAGAG 533
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 534 ATTCAATAGGCTCTGAAGCTGAAGACCGGTGGTCACTTACCTAGTTGTAATTCAAAAGTATT 593
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValValAspSerLysLeuAsp 200
Db 594 TACATGGCAAGAGGCTGTGCAGCTACCGGGTACTACTATGTTGATCTCCAAACTGGAT 653
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArg 220
Db 654 ATAAACAGCCCAAGAGACTATACAAATCGTTGAGCAGTATGAAGAACCAGGAGGAGCGC 713
Qy 221 HisHisLeuPheLeu 225
Db 714 CACCATCTGTTCTCTT
```

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 484 ATCCACAGGCGCCCTGAAGCTGAAGGACGGCGGCACCTACTCTGGTGGAGTTCAAGTCCATC 543
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 544 TACATGCGCAAGAGCCGTCAGCTGCCGGCTACTACTACGTGGACTCCAGCTGGAC 603
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 604 ATCACCAGCCCAACAGGAGCTACACCATCGTGAGCAGTACGAGAGGACCGAGGGCAGG 663
Qy 221 HisHisLeuPheLeu 225
Db 664 CACCACCTGTTCTCTG 678
RESULT 10
AX233627
LOCUS AX233627 713 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 51 from Patent WO0162919.
ACCESSION AX233627
VERSION AX233627.1 GI:15593330
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Nelson, D., Zamaira, E. and Tsien, R.
TITLE Modified fluorescent proteins
JOURNAL Patent: WO 0162919-A 51 30-AUG-2001;
Aurora Biosciences Corporation (US)
FEATURES
Location/Qualifiers
1..713
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Mutant Anthozoan red fluorescent protein"
ORIGIN
Alignment Scores:
Pred. No.: 7,41e-119 Length: 713
Score: 1210.00 Matches: 224
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 99.67% Indels: 0
DB: 6 Gaps: 0
US-10-006-922A-12 (1-225) x AX233627 (1-713)
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 22 GTGAGGAGCAGCAAGACGTGATCAAGGAGTTTCATGAGTTCAAGGTTCCAGGTCGCGATGGAGGC 81
Qy 21 ThrValasnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 82 ACCGTGAACGCCACGAGTTTCAGATCGAGGCGGAGGCGGAGGCGCGCCCTACGAGGC 141
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTtpAspile 60
Db 142 CACAAACCGTGAAGCTTAAGTGACCAAGGCGGCGCCCTCGCCCTTCGCTGGGACATC 201
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspilePro 80
Db 202 CTGAGCCCCAGTTCCAGTACGCGCAGCAAGGTGACGTGAAGCACCCCGCGCATCCCC 261
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnGlu 100
Db 262 GACTACAGAGCTGAGCTTCCCGGAGGCTTCAAGTGGGAGGAGGTGATGAATTCGAG 321
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 322 GACGGCGCGTGGTGACCGGTGACCCAGACAGCAGCTGCGAGGACGGCTCTTCATCTAC 381

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 382 ARGGTGAAGTTCATCGCGCTGAACCTTCCCGAGCGAGCGCCCGGTGATGCAAGAGACC 441
Qy 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 442 ATGGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGCGACGGCTGCTGAGGGCGAG 501
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 502 ATCCACAGGCGCTGAAGCTGAAGGACGGCGGCACCTACTCTGGTGGAGTTCAAGTCCATC 561
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 562 TACATGCGCCAAAGACCGCTGCAGCTGCCCGGTACTACTACTACGTGGACTCCAGCTGGAC 621
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 622 ATCACCAGCCCAACAGGAGCTACACCATCGTGAGCAGTACGAGAGGACCGAGGGCAGG 681
Qy 221 HisHisLeuPheLeu 225
Db 682 CACCACCTGTTCTCTG 696
RESULT 11
AR527331
LOCUS AR527331 723 bp DNA linear PAT 08-OCT-2004
DEFINITION Sequence 1 from patent US 6723537.
ACCESSION AR527331
VERSION AR527331.1 GI:53914309
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 723)
AUTHORS Peelle, B.
TITLE Directed evolution of protein in mammalian cells
JOURNAL Patent: US 6723537-A 1 20-APR-2004;
FEATURES
Location/Qualifiers
1..723
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 7,54e-119 Length: 723
Score: 1210.00 Matches: 224
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 99.67% Indels: 0
DB: 6 Gaps: 0
US-10-006-922A-12 (1-225) x AR527331 (1-723)
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 4 GTGCGCTCTCTCCAAAGAACGTCATCAAGGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 63
Qy 21 ThrValasnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 64 ACCGTGAACGCCACGAGTTTCAGATCGAGGCGGAGGCGGCGCCCTACGAGGC 123
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTtpAspile 60
Db 124 CACAAACCGTGAAGCTTAAGTGACCAAGGCGGCGCCCTCGCCCTTCGCTGGGACATC 183
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspilePro 80
Db 184 CTGTCCCCCAGTTCAGTACGCGCTCCAGGTGTCAGTGAAGCACCCCGCGCATCCCC 243
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnGlu 100

```

Db      244  GACTACAAGAGCTGTCTCTCCCGAGGGCTTCAAGTGGGAGCGCGTGAATCTCGAG 303
Qy      101  AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db      304  GACGGCGCGTGGTGAACCGTGAACCGAGACTCTCCCTGCGAGGAGCGGTCTTCATCTAC 363
Qy      121  LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db      364  AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGACGGCCCCCGTAATGCAAGAGACC 423
Qy      141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db      424  ATGGCGTGGAGGCGCTCCACCGAGCGCGTGTACCCCGCGACGCGGTGCTGAAGGGCGAG 483
Qy      161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db      484  ATCCACAAGGCCCTGAAGCTGAAGAGCGGCGGCACCTACCTGGTGGAGTTCAAGAGATC 543
Qy      181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db      544  TACATGCCAAGAGCCGTCGACGTGCCCGGCTACTACTACGTGGACTCCAAAGCTGGAC 603
Qy      201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db      604  ATCACCCTCCCAACAGGAGCTACACCATCTGTGGAGCAGTACGAGCGCAGCGAGGGCGCG 663
Qy      221  HisHisLeuPheLeu 225
Db      664  CACCACCTGTTCTCTG 678

RESULT 12
LOCUS      CO882115                2721 bp    DNA        linear    PAT 11-OCT-2004
DEFINITION Sequence 1 from Patent WO2004083445.
ACCESSION  CO882115
VERSION     CO882115.1  GI:54034825
KEYWORDS   .
SOURCE     synthetic construct
ORGANISM   synthetic construct
REFERENCE  1
AUTHORS    Chavancy,G., Couble,P., Durand,B., Grenier,A.M., Horard,B.,
           Julien,E., Mauchamp,B., Nony,P., Prudhomme,J.C. and Royer,C.
TITLE      Nucleic acid controlling the expression of a useful polypeptide in
           the posterior silk glands of a lepidoptera and application thereof
JOURNAL    Patent: WO 2004083445-A 1 30-SEP-2004;
           Centre National de La Recherche Scientifique-CNRS (FR); UNIVERSITE
           CLAUDE BERNARD - LYON 1 (FR); Institut National de la Recherche
           Agronomique (INRA) (FR)
FEATURES   Location/Qualifiers
            source             1..2721
                                /organism="synthetic construct"
                                /mol_type="unassigned DNA"
                                /db_xref="taxon:32630"
                                /note="Description de la s quence artificielle : s quence
                                de fusion"

ORIGIN
Alignment Scores:
Pred. No.:      4.01e-118      Length:      2721
Score:          1210.00      Matches:      224
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 99.56%      Mismatches: 0
Query Match:    99.67%      Indels:      0
DB:             6      Gaps:      0

US-10-006-922A-12 (1-225) x CO882115 (1-2721)

Qy      1  MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValAtcMetGluGly 20
Db      2044  GTGGCGTCTCTCAAGAACGTATCAAGAGATTCAAGCGCTTCAAGGGTGGCATGGAGGGC 2103
Qy      21  ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40

```

```

Db      2104  ACCGTGAACGGCCACCGAGTTCGAGATCGAGGGCGAGGGCGCGCCCTACGAGGC 2163
Qy      41  HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db      2164  CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCCCTTCGCCCTTGGGACATC 2223
Qy      61  LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db      2224  CTGTCCCCCGAGTTCAGTACGGCTCCAAGGTGTAGTGAAGCACCCTCCGCGCATCCCC 2283
Qy      81  AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db      2284  GACTACAAGAAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCTGATGAACCTTCGAG 2343
Qy      101  AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db      2344  GACGGCGGCGTGGTGAACCGTGAACCTTCCCTCCGAGCGGCTCTTCATCTAC 2403
Qy      121  LysValLysPheIleGlyValAsnPheProSerAspGlyGlyHisTyrLeuValGluPheLysThr 140
Db      2404  AAGTGAAGTTTCATCGGCGTGAACCTTCCCTCCGAGCGGCCCCGTATGCAAGAGAGACC 2463
Qy      141  MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db      2464  ATGGCGTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGACGCGCTGTGAAGGGCGAG 2523
Qy      161  IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db      2524  ATCCACAAGGCCCTGAAGCTGAAGAGCGGCGGCACCTACTCTGGTGGAGTTCAAGTCCATC 2583
Qy      181  TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db      2584  TACATGCCAAGAGCCGTCGACGTGCCCGGCTACTACTACTCTGGACTCCAAAGCTGGAC 2643
Qy      201  IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db      2644  ATCACCCTCCCAACAGGAGCTACACCATCTGTGGAGCAGTACGAGCGCAGCGAGGGCGCG 2703
Qy      221  HisHisLeuPheLeu 225
Db      2704  CACCACCTGTTCTCTG 2718

RESULT 13
LOCUS      CO882117                2772 bp    DNA        linear    PAT 11-OCT-2004
DEFINITION Sequence 3 from Patent WO2004083445.
ACCESSION  CO882117
VERSION     CO882117.1  GI:54034827
KEYWORDS   .
SOURCE     synthetic construct
ORGANISM   synthetic construct
REFERENCE  1
AUTHORS    Chavancy,G., Couble,P., Durand,B., Grenier,A.M., Horard,B.,
           Julien,E., Mauchamp,B., Nony,P., Prudhomme,J.C. and Royer,C.
TITLE      Nucleic acid controlling the expression of a useful polypeptide in
           the posterior silk glands of a lepidoptera and application thereof
JOURNAL    Patent: WO 2004083445-A 3 30-SEP-2004;
           Centre National de La Recherche Scientifique-CNRS (FR); UNIVERSITE
           CLAUDE BERNARD - LYON 1 (FR); Institut National de la Recherche
           Agronomique (INRA) (FR)
FEATURES   Location/Qualifiers
            source             1..2772
                                /organism="synthetic construct"
                                /mol_type="unassigned DNA"
                                /db_xref="taxon:32630"
                                /note="Description de la s quence artificielle : s quence
                                de fusion"

ORIGIN
Alignment Scores:
Pred. No.:      4.1e-118      Length:      2772

```

```
Score: 1210.00 Matches: 224
Percent Similarity: 100.00% Conservatives: 1
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 99.67% Indels: 0
DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x Q08B2117 (1-2772)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 2095 GTGGCGCTCTCCAAAGAACGTCATCAAGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 2154
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 2155 ACCGTGAACGCCACAGTTTCAGATCGAGGCGGAGGCGCGCCCTACGAGGCG 2214
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 2215 CACAACACCGTGAAGCTGAAGTGACCAAGGCGGCGCCCTGCCCTTCGCTTGGGACATC 2274
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 2275 CTGTCCCCCGAGTTCAGTACGGCTCCAAAGGTGACGTGAAGACACCCCGCGACATCCCC 2334
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnPheGlu 100
Db 2335 GACTACAAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAG 2394
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 2395 GACGGCGCGTGTGACCGTGACCGACGACGACTCTCTCCCTGCGAGGCGGTGTTTCATCTAC 2454
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 2455 AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGACGCGCCCGTAATCGAAGAAGACC 2514
Qy 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 2515 ATGGCGTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGACGCGGTGCTGAAGGCGAG 2574
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 2575 ATCCACAGGCGCTTGAAGCTGAGGAGCGCGGCGCACACTACCTGTGTGGAGTTCAAGTCCATC 2634
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 2635 TACATGGCCAAAGAGCGCGTGCAGTCCCGGCTACTACTACGTGGACTCCCAAGCTGGAC 2694
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 2695 ATCACCTCCCAACAGAGGACTACACCATCGTGAGCAGTACGAGCGCAGCGAGGCGCGC 2754
Qy 221 HisHisLeuPheLeu 225
Db 2755 CACCACCTGTTCCTG 2769

RESULT 14
LOCUS AX463702 4692 bp DNA linear PAT 15-JUL-2002
DEFINITION Sequence 16 from Patent WO0248338.
ACCESSION AX463702
VERSION AX463702.1 GI:21886461
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.

REFERENCE
1 Lichtenberg-Frat,H.
AUTHORS
TITLE Yeast strain for testing the geno- and cytotoxicity of complex
environmental contamination
JOURNAL Patent: WO 0248338-A 16 20-JUN-2002;
Lichtenberg-Frat, Hella (DE)
FEATURES
Location/Qualifiers

source 1. -4692
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/notes="Vektor pBere1-N1"

ORIGIN

Alignment Scores:
Pred. No.: 7.96e-118 Length: 4692
Score: 1210.00 Matches: 224
Percent Similarity: 100.00% Conservatives: 1
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 99.67% Indels: 0
DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x AX463702 (1-4692)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 682 GTGGCGCTCTCCAAAGAACGTCATCAAGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 741
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 742 ACCGTGAACGCCACAGTTTCAGATCGAGGCGGAGGCGCGCCCTACGAGGCG 801
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 802 CACAACACCGTGAAGCTGAAGTGACCAAGGCGGCGCCCTGCCCTTCGCTTGGGACATC 861
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 862 CTGTCCCCCGAGTTCAGTACGGCTCCAAAGGTGACGTGAAGACACCCCGCGACATCCCC 921
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnPheGlu 100
Db 922 GACTACAAAGAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAG 981
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 982 GACGGCGCGTGTGACCGTGACCGACGACGACTCTCTCCCTGCGAGGCGGTGTTTCATCTAC 1041
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 1042 AAGGTGAAGTTTCATCGGCGTGAATCTCCCTCCGACGCGCCCGTAATCGAAGAAGACC 1101
Qy 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 1102 ATGGCGTGGAGGCGCTCCACCGAGCGCTGTACCCCGCGACGCGGTGCTGAAGGCGAG 1161
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 1162 ATCCACAGGCGCTTGAAGCTGAAGGAGCGGCGGCGCACCTACCTGTGTGGAGTTCAAGTCCATC 1221
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 1222 TACATGGCCAAAGAGCGCGTGCAGTCCCGGCTACTACTACGTGGAGTCCCAAGCTGGAC 1281
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 1282 ATCACCTCCCAACAGAGGACTACACCATCGTGAGCAGTACGAGCGCAGCGAGGCGCGC 1341
Qy 221 HisHisLeuPheLeu 225
Db 1342 CACCACCTGTTCCTG 1356

RESULT 15
LOCUS AX823860 6893 bp DNA linear PAT 11-DEC-2003
DEFINITION Sequence 8 from Patent WO03070931.
ACCESSION AX823860
VERSION AX823860.1 GI:39750176
KEYWORDS
SOURCE
synthetic construct
```

```
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE
1
AUTHORS Baum,C., Will,E., Ostertag,W., Klump,H. and Schiedlmeier,B.
TITLE Methods for conducting site-specific dna recombination
JOURNAL Patent: WO 03070931-A 8 28-AUG-2003;
Vision 7 GmbH (DE)
FEATURES
Location/Qualifiers
source
1..6893
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Beschreibung der kuenstlichen Sequenz: provirale
Plasmid-DNA; retrovirales Cre-Reporterkonstrukt #SFR#"
misc_feature
1..435
/note="Plasmid-Rueckgrat (pUC)"
misc_feature
436..996
/note="MPSV 5'-LTR (Delta1-31)"
primer_bind
997..1014
/note="PBS (primer binding site)"
5'UTR
1015..1560
misc_feature
1561..1599
/note="#loxP1#-Sequenz"
misc_feature
1600..2289
/note="DsRed1-Gen aus Discosoma sp."
misc_feature
2290..2343
/note="#loxP2#-Sequenz"
misc_feature
2347..3150
/note="eGFP(tag)-Gen aus Aequorea victoria"
misc_feature
3164..4019
/note="#wPRE#"
3'UTR
4020..4080
misc_feature
4081..4641
/note="3'-LTR"
misc_feature
4642..6893
/note="Plasmid-Rueckgrat (pUC)"

ORIGIN
Alignment Scores:
Pred. No. : 1.29e-117 Length: 6893
Score: 1210.00 Matches: 224
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 0
Query Match: 99.67% Indels: 0
DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x AX823860 (1-6893)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db ::::
1603 GTGGCTCCTCCAAGACGTCATCAAGGAGTTCATGCGTTCAAGGTGGCATGGAGGGC 1662

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db ::::
1663 ACCGTGAACGGCCACGAGTTCGAGATCGAGGCGAGGGCGCGCCCTACGAGGGC 1722

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPheAlaTrpAspIle 60
Db ::::
1723 CACAACACCGTGAAGCTGAAGTGACCAAGGGCGGCCCCCTGCGCTTCGGGACATC 1782

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db ::::
1783 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCCGCGACATCCC 1842

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db ::::
1843 GACTACAAGAAGAGCTGTCCTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAG 1902

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db ::::
1903 GACGGCGCGGTGTGACCGTGACCGAGACTCTCCCTGCGAGGACGGCTGCTTCTATC 1962

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db ::::
2022 AAGTGAAGTTTCAATCGGCGTGAACCTCCCTCCGACGGCCCCCTAATGCAGAGAAGACC 2082
141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
2023 ATGGGCTGGGAGGCCCTCCACCGAGCGCTGTACCCCGGACGCGCTGTGAAGGGCGAG 2082
161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
2083 ATCCACAGGCCCTGAAGCTGAAGGAGCGGGGCCACTACTCTGGTGAGTTCAAGTCCATC 2142
181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
2143 TACATGCCCAAGAAGCCCGTGCAGCTGCGCGGCTACTACTACGTGGACTCCAAGCTGGAC 2202
201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
2203 ATCACCTCCACACACAGGAGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGC 2262
221 HisHisLeuPheLeu 225
2263 CACCACCTGTTCCTG 2277

Search completed: July 1, 2005, 19:55:32
Job time : 4404 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:39:59 ; Search time 40 Seconds
(without alignments)
541.219 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSSKNVKEFMFKVRMEG.....EDYIVQEYRTGRHHLFL 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	211.5	17.4	238	1 JQ1514	green-fluorescent
2	93.5	7.7	26926	1 I38344	titin, cardiac mus
3	92	7.6	1873	2 T30944	surface protein pr
4	91	7.5	458	1 A42386	hsp 90-binding pr
5	89	7.3	1433	1 A36734	bacillopeptidase F
6	87.5	7.2	1116	2 B70476	hypothetical prote
7	87	7.2	568	2 T06489	probable peptidylp
8	87	7.2	1484	2 C97196	probable membrane
9	85.5	7.0	456	2 G59397	signal-transducing
10	85	7.0	271	2 F69442	hypothetical prote
11	84.5	7.0	679	2 A40351	adhesion-type prot
12	84.5	7.0	725	1 IJMSNG	neural cell adhesi
13	84	6.9	1616	2 T17894	S-layer protein -
14	83.5	6.9	268	2 E30276	conserved hypothet
15	83.5	6.9	340	2 E89544	hypothetical prote
16	83.5	6.9	374	2 T06245	gibberellin 3 beta
17	83.5	6.9	374	2 T06244	gibberellin 3 beta
18	83	6.8	373	2 T50605	hypothetical prote
19	83	6.8	15281	2 S41309	cyclosporin synthe
20	82.5	6.8	559	1 S55383	peptidylprolyl iso
21	82.5	6.8	680	2 S37982	Kallmann syndrome
22	82.5	6.8	931	2 T32919	hypothetical prote
23	82	6.8	341	2 B53125	restriction enzyme
24	81.5	6.7	292	2 C59106	glucose-1-phosphat
25	81.5	6.7	296	2 I37989	La 4.1 protein - h
26	81.5	6.7	551	1 S72485	peptidylprolyl iso
27	81.5	6.7	629	2 C64180	hypothetical prote
28	81.5	6.7	6805	2 S20901	titin - rabbit (fr
29	81	6.7	346	2 S77025	nitrilase (EC 3.5.

30 81 6.7 587 1 E69171 phosphoesterase-re
31 80.5 6.6 404 1 S03849 ribonucleoprotein
32 80.5 6.6 415 1 JC1494 ribonucleoprotein
33 80.5 6.6 862 2 F75116 hypothetical prote
34 80.5 6.6 1003 1 AJCHPR phosphoribosylamin
35 80.5 6.6 1115 1 IJMSNL neural cell adhesi
36 80 6.6 621 2 A95250 choline binding pr
37 80 6.6 690 2 F98114 choline-binding pr
38 79.5 6.5 459 2 A46372 immunophilin FKBP5
39 79.5 6.5 475 2 T44566 conserved hypothet
40 79.5 6.5 487 2 T45982 hypothetical prote
41 79.5 6.5 648 2 B84139 ABC transporter (p
42 79.5 6.5 853 1 IJBONC neural cell adhesi
43 79 6.5 458 2 JN0873 immunophilin p59 -
44 79 6.5 703 2 A64351 hypothetical prote
45 79 6.5 1214 2 T47438 disease resistance

ALIGNMENTS

RESULT 1

JQ1514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text change 09-Jul-2004

C:Accession: JS0692; JQ1514; PQ0335; S48693; S51330; S51331

R:Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.

Gene 111, 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: JQ1514; MUID:92175527; PMID:11347277

A:Accession: JS0692

A:Molecule type: DNA

A:Residues: 1-107,'S',109-238 <PRA1>

A:Cross-references: UNIPROT:P42212; UNIPROT:Q17106; UNIPROT:Q17105; GB:M62654; NID:g1556

A:Accession: JQ1514

A:Molecule type: mRNA

A:Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>

A:Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661

A:Accession: PQ0335

A:Molecule type: protein

A:Residues: 46-64;74-122,132-151;154-183;185-200 <PRA3>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>

A:Cross-references: GB:L29345; NID:g606383; PIDN:AAA58246.1; PID:g606384

R:Watkins, J.N.; Campbell, A.K.

submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Accession: S51330

A:Molecule type: mRNA

A:Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'R',

A:Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009

A:Experimental source: clone gfp1

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24,'Q',26-29,'R',31-83,'L',85-153,'G',155-156,'P',158-171,'K',173-208,'Q',

A:Cross-references: EMBL:X83960; NID:g634010; PIDN:CAA58790.1; PID:g634011

A:Experimental source: clone gfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation; X-ray crystallography, 1.9 angstroms, residues 'A',2-79,'R',81-9

A:Note: engineered sequence based on JQ1514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

A;Contents: annotation; X-ray crystallography, 1.9 angstroms
C;Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
C;Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
C;Genetics:
A;Gene: GFP
A;Introns: 69/3; 167/3
C;Superfamily: green-fluorescent protein
C;Keywords: chromoprotein; luminescence
F;65-67/Cross-link: 5-imidazolone (Ser-Gly) #status experimental
F;66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 17.4%; Score 211.5; DB 1; Length 238;
Best Local Similarity 25.7%; Pred. No. 3.6e-11;
Matches 56; Conservative 47; Mismatches 102; Indels 13; Gaps 6;

Qy 16 VRMEGTVNGHEFELEGEGRPYEGHNTVKLVKTKGGPLPAWDILSPQYQSKVYVKH 75
Db 16 VELDGVNGHKFSGSGEGDATYKGLTPICTT-GKLPVPWPLTTFYSYGVQCFSRY 74
Qy 76 PADIP--DYKKLSPPEGFKWERVNVNFDGGVVTVTQDSSLQDGCIFYKVFIGNPSPDG 133
Db 75 PDHMKQHDFFKSNPEGVQERTIFYKDDGNKYTRAEVKPEGDTLVNRIELKGIKDFKEDG 134
Qy 134 PVMQKKTGWBEASTERLYPRDGLVKGRIHKKALKLK----DGGHYLVVEF--KSIYMAKKPV 187
Db 135 NILGHK-MEYNYNSHNVYINADKQNGIKVNFKIRHNIEDGVSQVLADHYQONTPIGDGPV 193
Qy 188 QLPGYVYVDSKLDIT---SHNEDYTVIEQYERTEGRH 222
Db 194 LLPDNHYLSQTQALSUKDPNEKRDMHILLEFVTAAGITH 231

RESULT 2

I38344
titin, cardiac muscle [validated] - human
N;Alternate names: connectin
N;Contents: serine/threonine-specific protein kinase (EC 2.7.1.1-)
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004
C;Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R;Label: S.; Kolmerer, B.
Science 270, 293-296, 1995
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A;Reference number: A57430; MUID:96026330; PMID:7569978
A;Accession: I38344
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A;Molecule type: mRNA
A;Residues: 1-26926 <LAB1>
A;Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425
R;Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A;Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A;Reference number: I38345; MUID:95119041; PMID:7819249
A;Accession: I38345
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1977-2014 <MUS>
A;Cross-references: EMBL:X83270; NID:9602579; PIDN:CAA58243.1; PID:g602580
A;Note: conformation and properties are reported for a synthetic peptide corresponding t
R;Label: S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380; PMID:1582406
A;Accession: S20898
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 13597-14200, 'I', 14202-14696 <LAB2>
A;Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A;Accession: S20897
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A;Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191

A;Accession: S20899
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-22484
A;Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195
R;Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A;Title: Genomic organization of M line titin and its tissue-specific expression in two c
A;Reference number: S63665; MUID:96177761; PMID:8604138
A;Accession: S63665
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 26729-26825 <KOL>
A;Cross-references: EMBL:X92412; NID:g1236761
R;Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A;Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentiat
A;Reference number: S37393; MUID:94008990; PMID:8404852
A;Accession: S37393
A;Molecule type: mRNA
A;Residues: 26831-26926 <GAU>
R;Improt, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A;Reference number: A66736; PDB:1TIT.
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R;Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A66201; PDB:1NCT
A;Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q32
C;Function:
A;Description: structural protein forming filaments in striated muscle
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prot
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein
F;24752-25008/Domain: protein kinase homology <KIN>
F;84,177,905,2276,2378,2459,2481,2563,2569,2763,2896,3088,3179,3384,3432,3628,3772,4068,41
98,11066,11486,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
F;16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
F;1900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,2484
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 7.7%; Score 93.5; DB 1; Length 26926;
Best Local Similarity 20.7%; Pred. No. 4.2e+02;
Matches 51; Conservative 34; Mismatches 84; Indels 77; Gaps 13;

Qy 15 KVRMEGTVNGHEFE-----IEGEGGRPYEGHNTV-----KLKVTKGPL 54
Db 23493 RYKSTGLTEGLEHYHRVTAINARGSKRPSKPIVAMDPIAPGKPNRVTDTTTSV 23552
Qy 55 PFAWDILSPQYQSKV--YV-----KH-----PADIPDYKKLSPPEGFKWE-RV 96
Db 23553 SLASV--PEDEGSKVTGYLIEMQVQDHEWTKCNTPTTKIREYTLTLPLPQCAEYFRV 23610
Qy 97 MNPEDG-----GVVVTQ-----DSSLQDGCIFYKVKFIGNVFPDSDG---PV 135
Db 23611 LACNAGGPGPEAEVPGTVKVTLEMPDYELDERYQEGIFVRQGVIRLTPIKGFPPFI 23670
Qy 136 MQKKTGWBEASTERLYPRDGLVKGRIHKKALKLK----DGGHYLVVEF-----KSIYMAK 184
Db 23671 CK-----WTKEGQDISKRAMIATSTETHLTIVTKADRGDSGTVDLVLENKCGKAVIKV 23725
Qy 185 KPVQLP 190
Db 23726 RVIGSP 23731

RESULT 3
T30944

surface protein precursor - Enterococcus faecalis

C;Species: Enterococcus faecalis
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30944
R;Shankar, V.; Baghdavan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.
Infect. Immun. 67, 193-200, 1999
A;Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene encoding surface protein precursor
A;Reference number: 220943; MUID:99081742; PMID:9864215
A;Accession: T30944
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1873 <SHA>
A;Cross-references: UNIPROT:Q924N7; EMBL:AF034779; NID:G3873187; PIDN:AA00

Query Match 7.6%; Score 92; DB 2; Length 1873;
Best Local Similarity 20.8%; Pred. No. 17;
Matches 46; Conservative 33; Mismatches 56; Indels 86; Gaps 13;
QY 30 EGECEGRPYEGHNTVKKVTKGGPLPFAWDILSPQFOYGS-----KVVVKHPA 77
DB 708 KGVGESEPTIG--TIPIKIV-----YQGSVGTDTDLAVTSKNIYENPGE 750
QY 78 DIP--DYKKLSPPECFKWRVNFEDGGVVTVDQSSLDQGCIFYKVFIGNPSPDG--PV 135
DB 751 NIPAGYHKVYFTAG-----EGTSIESGTTVPVK--DGVSLPEDKLPV 791
QY 136 MQKMTMGWEASTERLYPRDGV--LKGE-----IHKALKLKD-----GGHYLVF-- 177
DB 792 LKAK-----DGVTDKAPPEATQPIKADDTFVSSATKLDIIENPGDNIPAGYHKVYFTFA 847
QY 178 -----KSIYNAKVPQLPGYVYVDSKLDITSHNEDYT 209
DB 848 GEGTSIESGTTVFAKGVSLP-----EDKLPVLKAKDGYT 883

RESULT 4

A42386
hsp 90-binding protein p59 - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A42386
R;Lebeau, M.C.; Massol, N.; Herrick, J.; Faber, L.E.; Renoir, J.M.; Radanyi, C.; Baulieu
J. Biol. Chem. 267, 4281-4284, 1992
A;Title: Chem. 267, 4281-4284, 1992
A;Reference number: A42386; MUID:92165768; PMID:1537818
A;Accession: A42386
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-458 <LEB>
A;Cross-references: UNIPROT:P27124; GB:M84988; NID:G165594; PIDN:AAA31439.1; PID:G165595
A;Experimental source: liver
A;Note: sequence extracted from NCBI backbone (NCBIP:83839)
C;Superfamily: human FK506-binding protein FKBP51; BKBP-type peptidylprolyl isomerase h
F;50-97/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F;319-352/Domain: tetratricopeptide repeat homology <TPR>
F;353-386/Domain: tetratricopeptide repeat homology <TPR>

Query Match 7.5%; Score 91; DB 1; Length 458;
Best Local Similarity 24.6%; Pred. No. 3.2;
Matches 42; Conservative 26; Mismatches 49; Indels 54; Gaps 10;
QY 57 AWDI-----LSPQFOYGSKVYVYKHPADIPDYKLSFPE---GFKWRVYMF 99
DB 89 AWDIAVATMKVGELCRITCKPEYAGS---AGSPKIPPNATLVFEVLEFEGKEDLTDD 145
QY 100 EDGGVVTVDQSSLDQGCIFYKVFIGNP--PDGPMQKMTGWASTERLYP----R 153
DB 146 EDGGI-----IRRIIRGEVYARPNDGAIVALEGY---YKDRFDQREL 189
QY 154 DGVLLKGE-----IHKAL-KLDGGHYLVFEEKSIY---MAKPKVQLPGY 192
DB 190 FEVGESELDLPCLGLEKAIQRMKEGHSILYKPSYAFGNAGREKFPQIPPY 240

RESULT 5

A36734
bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 09-Jul-2004
C;Accession: A36734; A35131; A35750; B35750; S08223; JN0335; I39849; B69596; JU0084
R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 5520-5521, 1990
A;Reference number: A36734; MUID:90368623; PMID:2118514
A;Contents: erratum
A;Accession: A36734
A;Molecule type: DNA
A;Residues: 1-1433 <SLO>
A;Cross-references: UNIPROT:P16397; GB:M29035; NID:G143307; PIDN:AAA62679.1; PID:G143308
R;Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 1470-1477, 1990
A;Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning
A;Reference number: A35131; MUID:90170864; PMID:2106512
A;Accession: A35131
A;Molecule type: DNA
A;Residues: 1-365, 'S', 367-682, 'EIMP', 893, 'Q', 895-896 <SL2>
A;Cross-references: GB:M29035
A;Note: the authors translated the codon GAA for residue 545 as Leu
R;Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J. Biol. Chem. 265, 6845-6850, 1990
A;Title: Cloning, genetic organization, and characterization of a structural gene encoding
A;Reference number: A35750; MUID:90216713; PMID:2108961
A;Accession: A35750
A;Molecule type: DNA
A;Residues: 1-392, 'V', 394-828, 'NIRTRYLSLKFCRSRHKSV' <WUA>
A;Cross-references: GB:J05400; NID:G142607; PIDN:AAA83362.1; PID:G142609
A;Note: this sequence has been corrected
A;Accession: B35750
A;Molecule type: DNA
A;Residues: 876-935, 'CG', <WU2>
A;Cross-references: GB:J05400; NID:G142607; PIDN:AAA83363.1; PID:G1119197
R;Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A;Title: Nucleotide sequence of the sporulation gene spoIGA from Bacillus subtilis.
A;Reference number: S08223; MUID:90174995; PMID:2106671
A;Accession: S08223
A;Molecule type: DNA
A;Residues: 1410-1433 <MAS>
A;Cross-references: EMBL:X17344; NID:G40165; PIDN:CAA35224.1; PID:G809661
R;Kato, T.; Yamagata, Y.; Arai, T.; Ichihama, E.
Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
A;Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric p
A;Reference number: JN0335; MUID:93005071; PMID:1368833
A;Accession: JN0335
A;Molecule type: protein
A;Residues: 195-218, 'A' <KAT>
A;Note: source of this material was Bacillus subtilis (natto)
R;Seall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A;Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli
A;Reference number: I39846; MUID:89008108; PMID:3139638
A;Accession: I39849
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-211 <RES>
A;Cross-references: GB:M22630; NID:G142938; PIDN:AAA22458.1; PID:G551705
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsberg, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lardinois, A.; Lardinois
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serod
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69596
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1433 <KUN>
A;Cross-references: GB:Z99111; GB:Z99112; GB:AL009126; NID:G2633902; PIDN:CAB13404.1; PT
A;Experimental source: strain 168
C;Genetics:
A;Gene: bpr; bpf
A;Map position: 135 (degrees)
C;Superfamily: bacillopeptidase F; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-194/Domain: propeptide #status predicted <PRO>
F;195-1433/Product: bacillopeptidase F #status experimental <MAT>
F;218-466/Domain: subtilisin homology <SRT>
F;227,274,452/Active site: Asp, His, Ser #status predicted

Query Match 7.3%; Score 89; DB 1; Length 1433;
Best Local Similarity 24.0%; Pred. No. 22;
Matches 56; Conservative 26; Mismatches 71; Indels 80; Gaps 14;

Qy 17 RMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKVGGLPFA----- 57
Db 521 KAEQVS-----VEGDQEPVYQHEKV-TEAYEGSLPTLTAEADNVSVTSVKLSYKLD 574
Qy 58 ---WDILSPQFQGVKVVVHPADIPDYK--KLSFPEGFKWERVMNPFEDG-----VVT 106
Db 575 QGEWTEITAKRISDHLKGYQAEIPDIKTKLS----YKW--MIHDFGHVSSDVYD 627
Qy 107 VTQSSLODGCFFIYKVKF-----IGVN-----FPSDGPVMQKKTGWGEASTERLY 151
Db 628 VTVKPSITAG---YKQDFETAPGGWASGTTNNWEGVPSTGP-----NTAASGEKVY 677
Qy 152 PRDGVKGE-----IHKALKDKGGHYLVFEFKSIYMAKKPVQLPGYVYV 195
Db 678 GTN--LTGNVANSANMNLVMPPIKAPDSGSLFLQFKSWHNLDDFDY-GYVYV 727

RESULT 6
B70476
hypothetical protein aq_2054 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70476
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1116 <AAQF>
A;Cross-references: UNIPROT:O67838; GB:AE000770; NID:G2984274; PIDN:AAC07805.1; PID:G298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_2054

Query Match 7.2%; Score 87.5; DB 2; Length 1116;
Best Local Similarity 22.0%; Pred. No. 21;
Matches 56; Conservative 35; Mismatches 103; Indels 61; Gaps 9;

Qy 5 KNVKEFMRFKVRVEGTVNGHEFEIEGEGE-----GRPYEGHNTVKLVKTKGG 52
Db 164 KOLLDDSEYSAKTKGEIKRNTGTGELAEVEIKETKENFTLSGKYNKAGTINLPVLDIN 223
Qy 53 PLPPFANDILSPQFQVGS-----KVYVVKHPADIPDYKLSF-----PEG---F 91

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serod
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: B69596
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1433 <KUN>
A;Cross-references: GB:Z99111; GB:Z99112; GB:AL009126; NID:G2633902; PIDN:CAB13404.1; PT
A;Experimental source: strain 168
C;Genetics:
A;Gene: bpr; bpf
A;Map position: 135 (degrees)
C;Superfamily: bacillopeptidase F; subtilisin homology
C;Keywords: extracellular protein; hydrolase; serine proteinase
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-194/Domain: propeptide #status predicted <PRO>
F;195-1433/Product: bacillopeptidase F #status experimental <MAT>
F;218-466/Domain: subtilisin homology <SRT>
F;227,274,452/Active site: Asp, His, Ser #status predicted

Query Match 7.3%; Score 89; DB 1; Length 1433;
Best Local Similarity 24.0%; Pred. No. 22;
Matches 56; Conservative 26; Mismatches 71; Indels 80; Gaps 14;

Qy 17 RMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKVGGLPFA----- 57
Db 521 KAEQVS-----VEGDQEPVYQHEKV-TEAYEGSLPTLTAEADNVSVTSVKLSYKLD 574
Qy 58 ---WDILSPQFQGVKVVVHPADIPDYK--KLSFPEGFKWERVMNPFEDG-----VVT 106
Db 575 QGEWTEITAKRISDHLKGYQAEIPDIKTKLS----YKW--MIHDFGHVSSDVYD 627
Qy 107 VTQSSLODGCFFIYKVKF-----IGVN-----FPSDGPVMQKKTGWGEASTERLY 151
Db 628 VTVKPSITAG---YKQDFETAPGGWASGTTNNWEGVPSTGP-----NTAASGEKVY 677
Qy 152 PRDGVKGE-----IHKALKDKGGHYLVFEFKSIYMAKKPVQLPGYVYV 195
Db 678 GTN--LTGNVANSANMNLVMPPIKAPDSGSLFLQFKSWHNLDDFDY-GYVYV 727

RESULT 6
B70476
hypothetical protein aq_2054 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70476
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1116 <AAQF>
A;Cross-references: UNIPROT:O67838; GB:AE000770; NID:G2984274; PIDN:AAC07805.1; PID:G298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_2054

Query Match 7.2%; Score 87.5; DB 2; Length 1116;
Best Local Similarity 22.0%; Pred. No. 21;
Matches 56; Conservative 35; Mismatches 103; Indels 61; Gaps 9;

Qy 5 KNVKEFMRFKVRVEGTVNGHEFEIEGEGE-----GRPYEGHNTVKLVKTKGG 52
Db 164 KOLLDDSEYSAKTKGEIKRNTGTGELAEVEIKETKENFTLSGKYNKAGTINLPVLDIN 223
Qy 53 PLPPFANDILSPQFQVGS-----KVYVVKHPADIPDYKLSF-----PEG---F 91

Db 224 AKAFVADLVNRKNKYSGIEGVKNYVELFDKFLKGEAVNPECTKIKFTYDVIVPEGLLTF 283
Qy 92 KWER-VNWFEDGGVTVTQDSSLODGCFFIYKVKFICGVNFPSPDGPVMQKKTGWGEASTERL 150
Db 284 SFENLVVDKNTLGINREIRGEFHNGKGVDFKCMFVKV-----ATENL 327
Qy 151 YPRDGVKLGKEIHKALKDKGGHYLVFEFKSIYMAKKPVQLPGYVYV-D-SKLDITSHNEDYT 209
Db 328 EVIDKKFKGDVLPFSYNGSGSLNFEKSGYAK-----GNLIINKKLEGEFSFNDFF 381
Qy 210 IVEQ-----YERTEGR 220
Db 382 VVFQDFNAVLSGEGK 396

RESULT 7
T06489
probable peptidylprolyl isomerase (EC 5.2.1.8) FKBP77 - wheat
C;Species: Triticum aestivum (common wheat)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C;Accession: T06489
R;Brieman, A.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z15713
A;Accession: T06489
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-568 <BRI>
A;Cross-references: UNIPROT:O04843; EMBL:Y07636; PIDN:CAA68913.1
A;Experimental source: cv. ATIR, 2 day old plants, root tips
C;Genetics:
A;Gene: FKBP77
C;Superfamily: peptidylprolyl isomerase ROP1; BKBP-type peptidylprolyl isomerase homology
C;Keywords: calmodulin binding; cis-trans-isomerase; cyclosporin A binding
F;54-101/Domain: BKBP-type peptidylprolyl isomerase homology <PPI>
F;169-212/Domain: BKBP-type peptidylprolyl isomerase homology <PPI2>
F;286-334/Domain: BKBP-type peptidylprolyl isomerase homology <PPI1>
F;482-515/Domain: tetratricopeptide repeat homology <TPR>

Query Match 7.2%; Score 87; DB 2; Length 568;
Best Local Similarity 21.8%; Pred. No. 9.6;
Matches 61; Conservative 32; Mismatches 81; Indels 108; Gaps 15;

Qy 5 KNVKE-----FMRFKVRMEGT-VNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPL 54
Db 39 KLVKGEGWDTAETALKVEVHTGLDGTED-SSRDGTFF-----KFKLEGGQVI 91
Qy 55 PFAMD-----ILSPQFQVGSKVYVVKHPADIPDYKLSFP----- 88
Db 92 K-GWDQGIKTMKKGASLTIPDPLAYGERA----PRTIPPNTATLRFDVLLSWASVKOI 146
Qy 89 -----EGFKWERVMNPFEDGGVTVTQDSSLODGCFFIYKVKFICGVNFPSPDGPVM 136
Db 147 CKDGGIFKKLVVEGQKWE---NPKLDDEVTVKVEARLEDGSGVSVKSGESI----- 192
Qy 137 QKKTGWGEASTERLYPRDGVKLGKEIHKALK-LKDGCHYLVEFKSIY----MAKKPVQLPG 191
Db 193 -----EFSV-----KDGFCFALSKAVKTKMKKGKVLTVKPYQFCFGQGRAATEVEG 240
Qy 192 YYYVDSKLDI-----TSHNEDYTIV-----EQYER 216
Db 241 AVPPNSTLHIDLQVSWKTLTLIGDDKRILKVLKKEGEGYER 282

RESULT 8
C97196
probable membrane protein [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
C;Accession: C97196
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: C97196
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1484 <KUR>
A;Cross-references: UNIPROT:O97GG4; GB:AE001437; PIDN:AAK00358.1; PID:gl5025418; GSPDB:G
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2403

Query Match 7.2%; Score 87; DB 2; Length 1484;
Best Local Similarity 20.7%; Pred. No. 34;
Matches 62; Conservative 47; Mismatches 76; Indels 114; Gaps 17;

Qy 9 KEFMRFRVMEG-----TVNGHEPEIG-----EGEGRYEGHNTVKLVTKGGP 53
Db 766 KNIVDFKTLKNSFFVSENPREFKIEADIRKNSNAPEVKQIAEGDSANNWEVAKSGV 825

Qy 54 L-----PFAWDI-----LSPQFQ-----YGSKVYVHKHPADIPDYKLSF-- 87
Db 826 LKARENNPYKPKINISGKNTWKHLKAKFYDANMHEIYSS--YIVKPSDSVDPESMDFTG 883

Qy 88 ----PEGFKWER--VMNPE--GGVVTVTQDSSLDGCFYKVKFIGNFSPDGPVQMK 139
Db 884 ECVSPPKTKMRLEILSYKPKHKIYWKIHDVNIYD--LSEYKI-----SNSFVNKK 934

Qy 140 TMGWEASTERLYPRDGLVKGELHAKLKDGGHYLV-----EFKSIYM 182
Db 935 VK--KATAKYIR-----ALVSKGGKLVTVGKSVVTVDFDESKSGFKWYL 982

Qy 183 AKPVLPGYVYDS-----KLD-----ITSHNEDYTVBQYERTEGRHHLF 224
Db 983 GK-----FHPDSGKNSIKVENKQGFNAINLFTVIPENEYNTVTKKIEKTDKCNIF 1033

RESULT 9
G69397
signal-transducing histidine kinase homolog - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: G69397
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: G69397
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-456 <KLE>
A;Cross-references: UNIPROT:O29083; GB:AE001022; GB:AE000782; NID:G2689345; PIDN:AAB9005

Query Match 7.0%; Score 85.5; DB 2; Length 456;
Best Local Similarity 22.0%; Pred. No. 9.7;
Matches 46; Conservative 25; Mismatches 59; Indels 79; Gaps 10;

Qy 26 EFEIEGEGRPYEGHNTVKLVTKGGPLPFAWDILSPQFQYGSKVYVHKHPADIPDYK 85
Db 67 EFEEELGE-EGRFYSR-----YITKGRVVRHVWGFTAKFTPLGKSYIIGNWIDVTSKEL 120

Qy 86 SFPSPGFKWERVNFEDGGVTVTQDS-----SLDGGCFIY-----KVKFTGVN 128
Db 121 -----EQALRESEEFYRTLVDSLTVPVLLQDGMVYVYNNKAFEEATGYKREEIVGRN 172

Qy 129 -----FPSDGPVMOKK-----TMGWEASTERLYPRDGLV----- 158
Db 173 PFFLIHPEDRLGLVYKRIERKGLRDTMETYSW-----RIIRKDGVRVWTARPGRTYR 227

Qy 159 -----GEIHKA-LKLKGGHYL 174
Db 228 GRPAVAATVVYDTTEIHKLNIELKXKGEYL 256

RESULT 10
F69442
hypothetical protein AF1543 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: F69442
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
A;Reference number: A69250; MUID:98049343; PMID:9389475
A;Accession: F69442
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-271 <KLE>
A;Cross-references: UNIPROT:O28729; GB:AE000996; GB:AE000782; NID:G2689319; PIDN:AAB8971
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1420

Query Match 7.0%; Score 85; DB 2; Length 271;
Best Local Similarity 25.6%; Pred. No. 5.4;
Matches 44; Conservative 18; Mismatches 64; Indels 46; Gaps 9;

Qy 64 QFYGSKVYVHKP-----ADIPDYK--LSFPSPGFKWERVNFEDGGV---TVTQSSL 113
Db 65 QTEINSIFYTLHSLNLLGEDLPDYSSFLKRLKLEGLKAERKYLSDGGVTATYTFQPNAL 124

Qy 114 QDGFYVYVYVFGVNPSPDGPVQKTKMGWEASTERLYPRD-----GVLLKGIHKALK 166
Db 125 RDA--YMI STLHLNRYDVPETKMLV-----RRYRRETEFGVGVKPKNLEETFY 174

Qy 167 LKDGHYLVFEKSIY-----MAKPVQLPGY-----YVVDKSLDITS 203
Db 175 ----ASYILRDKAVISFVKSFESNGGFAKQGGYPPYLEDYVATSTLSLLS 222

RESULT 11
A40351
adhesion-type protein ADMLX - human
C;Species: Homo sapiens (man)
C;Date: 24-Jan-1992 #sequence_revision 24-Jan-1992 #text_change 15-Mar-2004
C;Accession: A40351
R;Legouis, R.; Hardelin, J.P.; Levilliers, J.; Claverie, J.M.; Compain, S.; Wunderle, V.
G.; Weissbach, J.; Petit, C.
Cell 67, 423-435, 1991
A;Title: The candidate gene for the X-linked Kallmann syndrome encodes a protein related
A;Reference number: A40351; MUID:92005720; PMID:1913827
A;Accession: A40351
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-679 <LEG>
A;Cross-references: GB:S60085; NID:G237595; PIDN:AAB20108.1; PID:G237597
C;Genetics:
A;Gene: GDB:KAL1; KAL
A;Cross-references: GDB:120116; OMIM:308700
A;Map position: Xp22.32-Xp22.32
P;129-175/Domain: antileukoprotease repeat homology <ALP>

Query Match 7.0%; Score 84.5; DB 2; Length 679;
Best Local Similarity 27.0%; Pred. No. 20;
Matches 48; Conservative 20; Mismatches 75; Indels 35; Gaps 10;

Qy 46 LKVTGKG---PLPF-----AWDILSPQFQYGS---KVVYVHKPAD--IPDYKLSPEG 90
Db 406 VKTRKGIQQLPQRRRPTRRPLEVGAPFYQDGLQVYVWKKTEDPTVRYHVRWPFPEA 465

```
Qy 91 FKWERVMNFDGGVYVTVTQDSSLQDGCFFIYKVKFIGNFSPDGPVMQKTMGWAEASTERL 150
Db 466 CAHNRRTTGSSEASSGTHENYIILQDLSFSCYKV-----TVQPIRPKS-----HSAEAV 515

Qy 151 Y----PRDGLVKGETHKALK-LKDGCHVLVEPKSIYMAKKVPQLPGYGYVDSKLDITSH 204
Db 516 FFTTPPCSAKLGKSHKPGICGLGEAGHVLSK-----VLAKPENTLSASFIVQD-VNITGH 567

RESULT 12
IJMSNG
Neutral cell adhesion molecule 1 precursor, GPI-anchored splice form - mouse
N/Alternate names: NCAM-120
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C/Accession: A29673; S00382; A44290
R/Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.
R/Barthele, D.; Santoni, M.J.; Wille, W.; Ruppert, C.; Chaix, J.C.; Hirsch, M.R.; Fontec
EMBO J. 6, 907-914, 1987
A/Title: Isolation and nucleotide sequence of mouse NCAM cDNA that codes for a Mr 79,000
A/Reference number: A29673; MUID:87246524; PMID:3595563
A/Accession: A29673
A/Molecule type: mRNA
A/Residues: 1-725 <BA>
A/Cross-references: UNIPROT:P13594; EMBL:Y00051; NID:G53342; PIDN:CAA68263.1; PID:G53343
R/Barbas, J.A.; Chaix, J.C.; Steinmetz, M.; Goridis, C.
EMBO J. 7, 625-632, 1988
A/Title: Differential splicing and alternative polyadenylation generates distinct NCAM t
A/Reference number: S00382; MUID:88283628; PMID:3396534
A/Accession: S00382
A/Molecule type: DNA
A/Residues: 642-656, 'D', 658-725 <BA2>
A/Cross-references: EMBL:X07195
R/Rougon, G.; Marehak, D.R.
J. Biol. Chem. 261, 3396-3401, 1986
A/Title: Structural and immunological characterization of the amino-terminal domain of m
A/Reference number: A44290; MUID:86140120; PMID:3512556
A/Accession: A44290
A/Molecule type: Protein
A/Residues: 20-36 <ROU>
C/Comment: NCAM mediates cell-cell adhesion via homophilic binding with another NCAM mol
C/Comment: Several forms of NCAM are produced by alternative splicing. See also PIR:IJMS
C/Genetics:
A/Gene: NCAM
A/Map position: 9
A/Introns: 701/1
C/Superfamily: neural cell adhesion molecule; fibronectin type III repeat homology; immu
C/Keywords: alternative splicing; cell adhesion; duplication; heparin binding; membrane
F/1-19/Domain: signal sequence #status predicted <SIG>
F/34-98/Domain: immunoglobulin homology <IMM1>
F/132-191/Domain: immunoglobulin homology <IMM2>
F/152-156/Region: heparin binding #status predicted
F/161-165/Region: heparin binding #status predicted
F/228-290/Domain: immunoglobulin homology <IMM3>
F/263-272/Region: NCAM binding #status predicted
F/323-388/Domain: immunoglobulin homology <IMM4>
F/420-482/Domain: immunoglobulin homology <IMM5>
F/519-596/Domain: fibronectin type III repeat homology <FN3A>
F/625-685/Domain: fibronectin type III repeat homology <FN3B>
F/41-96,139-189,235-288,330-386,427-480/Disulfide bonds: #status predicted
F/222,316,348,424,450,479/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 7.0%; Score 84.5; DB 1; Length 725;
Best Local Similarity 24.2%; Pred. No. 22;
Matches 37; Conservative 30; Mismatches 57; Indels 29; Gaps 9;

Qy 91 FKWERVMNFDGGVYVTV---TQDSSLQDGCFFIYKVKFIGNFSPDGPVMQKTMGWAEAST 147
Db 548 FTWYDAKEANMEGIVTMGLKPETYS-----RLAALNGKGLGEIMQPSSEKTPVP 600

Qy 148 ERLVPRDGLVKGETHK-----ALKLKG-----HYLVEPKSIYMAKKP-VQLP-GYY 193
Db 601 ELSAPK---LEGQMGDGNISKVNLIKQDDGGSPIRHYLVKYLALASEWKPEIRLPSGSH 657
```

```
Qy 194 YVDSK-LDITSHNEDYTIIVE-QYERTEGRHHLF 224
Db 658 HWMLKSLDWNNAEYEVYVAENQOQSKAAHFVP 690

RESULT 13
TI7884
S-layer protein - Bacillus circulans
C/Species: Bacillus circulans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: TI7884
R/Aubert-Fivert, E.; Davies, J.
Gene 147, 1-11, 1994
A/Title: Biosynthesis of butirosin in Bacillus circulans NRRL B3312: identification by se
A/Reference number: Z18808; MUID:94374689; PMID:7522196
A/Accession: TI7884
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-1616 <AUB>
A/Cross-references: UNIPROT:P35824; EMBL:L20421; NID:G304142; PID:G304143; PIDN:AAA62588
C/Genetics:
A/Gene: butB
C/Function:
A/Pathway: butirosin biosynthesis

Query Match 6.9%; Score 84; DB 2; Length 1616;
Best Local Similarity 20.8%; Pred. No. 70;
Matches 59; Conservative 48; Mismatches 85; Indels 94; Gaps 17;

Qy 6 NVIKEMR-----FKVRMEGTVNGHEFEI---EGEGEGRPYEGHNTVTKLVTK 50
Db 590 DLLQEFIRYSRELGLDIHVSNFNAEFSIASNEFALLDSDLHSDWEERYNAADNGQIK--- 646

Qy 51 GGLPFPANDILSPQFGSKVYVKHPAD-IPDYKKLSFFEGFKWERVMNFDGGVY--TV 107
Db 647 -----RLRESAKQGAFAFVNPNSDEVDRFQLKTI-----EVLQNYDVGWVLDRA 692

Qy 108 TDSSSLQDGCFFIYKVK---FIGV-----NFP-----SDGPVMO-----KKT 140
Db 693 RYNESADPSDLTKAFESFLGARGKQLQNPDDVTYAGNVKRDGLIRDWFEFRSKI 752

Qy 141 MGWEASTERLYPRDGLVKGELKALKL-----DGCHY-LVEPK---SIYMAK 184
Db 753 KSFTSEVRQLTDR---VKAEGKKIEVSAYVGSWFSESYLYLNGVHWSGTEFRYDERLRMKD 809

Qy 185 KPQLPGYY---YVDSKLDITSHNEDYTIIVEQYERT---EGRHHLFL 225
Db 810 KSVYTFGYBSSGYV-----KNLDFIMIGAYQTTAPEIEHYITL 847

RESULT 14
E90276
conserved hypothetical protein [imported] - Sulfolobus solfataricus
C/Species: Sulfolobus solfataricus
C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C/Accession: E90276
R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A/Description: Sulfolobus solfataricus complete genome.
A/Reference number: A99139
A/Accession: E90276
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-268 <KUR>
A/Cross-references: UNIPROT:Q97YT6; GB:AE006641; NID:G13814420; PIDN:AAK41468.1; GSPDB:GN
C/Genetics:
A/Gene: SS01221

Query Match 6.9%; Score 83.5; DB 2; Length 268;
Best Local Similarity 25.8%; Pred. No. 7.2;
Matches 49; Conservative 21; Mismatches 59; Indels 61; Gaps 13;
```

Search completed: June 30, 2005, 15:50:22
Job time : 42 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:31:48 ; Search time 175 Seconds
(without alignments)
658.388 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVKEFMRFKVRMEG.....EDYIVQEYTERGHHFL.225

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 03.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	1214	100.0	225	2 Q9U6Y8	Q9U6Y8 discosoma s
2	1196	98.5	236	2 Q6ND7	Q6nd7 discosoma s
3	1193	98.3	225	2 Q6F85	Q6f85 discosoma s
4	1191	98.1	236	2 Q6ND8	Q6nd8 discosoma s
5	1085.5	89.4	230	2 Q9GTU7	Q9gtj7 discosoma s
6	804	66.2	221	2 Q66PV1	Q66pv1 acropora te
7	798	65.7	221	2 Q66PV8	Q66pv8 acropora ac
8	797	65.7	221	2 Q66PV0	Q66pv0 acropora mi
9	793	65.3	221	2 Q95P04	Q95p04 gonopora t
10	793	65.3	221	2 Q66PU9	Q66pu9 acropora hy
11	751	61.9	219	2 P83690	P83690 montipora e
12	745	61.4	227	2 Q66ND6	Q66nd6 discosoma s
13	729.5	60.1	232	2 Q9U6T7	Q9uey7 discosoma s
14	690	56.8	225	2 Q963F5	Q963f5 montastraea
15	684	56.3	225	2 Q66ND3	Q66nd3 montastraea
16	681	56.1	225	2 Q95UA7	Q95ua7 montastraea
17	681	56.1	225	2 Q7Z0W4	Q7z0w4 montastraea
18	679.5	56.0	227	2 Q7Z0W6	Q7z0w6 montastraea
19	678.5	55.9	227	2 Q62E29	Q62ep9 montastraea
20	678.5	55.9	227	2 Q7Z0W8	Q7z0w8 montastraea
21	678	55.8	225	2 Q7Z0W5	Q7z0w5 montastraea
22	672.5	55.4	227	2 Q66ND2	Q66nd2 montastraea
23	670.5	55.2	227	2 Q66ND5	Q66nd5 montastraea
24	669.5	55.1	234	2 Q7Z0W7	Q7z0w7 montastraea
25	668.5	55.1	234	2 Q8T5F2	Q8t5f2 montastraea
26	667	54.9	266	2 Q9U6Y3	Q9uey3 clavularia
27	663	54.6	226	2 Q8T6U0	Q8t6u0 dendronspht
28	660.5	54.4	227	2 Q66ND4	Q66nd4 montastraea
29	655.5	54.0	226	2 Q66PU5	Q66pu5 agaricia fr
30	655.5	54.0	227	2 Q95V70	Q95vt0 montastraea
31	654.5	53.9	234	2 Q8MU47	Q8mu47 montastraea

ALIGNMENTS

RESULT 1

Q9U6Y8 PRELIMINARY; PRT; 225 AA.
AC Q9U6Y8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fluorescent protein FP583.
OS Discosoma sp.
OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=86600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Pradkov A.P., Labas Y.A., Savitsky A.P., Zarskiy A.G.,
RA Markelov M.L., Lukanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Matz M.V., Pradkov A.P., Labas Y.A., Savitsky A.P., Zarskiy A.G.,
RA Markelov M.L., Lukanov S.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF168419; AAF03369.1; -
DR PDB; 1G7K; X-ray; A/B/C/D=1-225.
DR PDB; 1GGX; X-ray; A/B/C/D=1-225.
DR GO; GO:0006091; p-energy pathways; IEA.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP-related.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 225 AA; 25931 MW; FBF9A5369778F689 CRC64;

Query Match 100.0%; Score 1214; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 2.9e-98;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
Qy 61 LSPQFOYGVSKVYKHPADIIDYKKLSPEGFKVRVNMFGDGGVTVTQSSLDGGCFIY 120
Db 61 LSPQFOYGVSKVYKHPADIIDYKKLSPEGFKVRVNMFGDGGVTVTQSSLDGGCFIY 120
Qy 121 KVFIFGVNFPDSGPFVMQKTMGWEASTERLYPRDGLVKGHIHKLKLDGGHYLVFVKSI 180
Db 121 KVFIFGVNFPDSGPFVMQKTMGWEASTERLYPRDGLVKGHIHKLKLDGGHYLVFVKSI 180
Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTYVEQVTEGRHHLPL 225

Db 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225

RESULT 2

Q66ND7 Q66ND7 PRELIMINARY; PRT; 236 AA.
AC Q66ND7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 23-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Enhanced red fluorescent protein R+.
OS Discosoma sp. RC-2004.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=289055;
RN [1]
RP SEQUENCE FROM N.A.
RA Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT "Cloning of Anthozoan Fluorescent Protein Genes";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY679107; AAU04444.1; -;
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR Pfam; PF01353; GFP; 1.
DR PRINTS: PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 236 AA; 27032 MW; BB3844BB06829EF0 CRC64;

Query Match 98.5%; Score 1196; DB 2; Length 236;
Best Local Similarity 98.2%; Pred. No. 1.2e-96;
Matches 221; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MSCSKNVIKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Qy 61 LSPQFQYGSKVYVVKHPADIPDYKLSFPEGFKWERVMNFEDGGVTVTQDSSLDGCFY 120
Db 61 LSPQFQYGSKVYVVKHPADIPDYKLSFPEGFKWERVMNFEDGGVTVTQDSSLDGCFY 120
Qy 121 KVFIGNVFPSPDGPVMQKTMGWEASTERLYPRDGLVKGEIHKALKLKDGGHYLVEPKSI 180
Db 121 KVFIGNVFPSPDGPVMQKTMGWEASTERLYPRDGLVKGEIHKALKLKDGGHYLVEPKSI 180
Qy 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225
Db 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225

RESULT 3

Q6KF85 Q6KF85 PRELIMINARY; PRT; 225 AA.
AC Q6KF85;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Orange fluorescent protein FP586.
OS Discosoma sp. JW-2002.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=208461;
RN [1]
RP SEQUENCE FROM N.A.
RA Wiedenmann J., Girod A.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF545828; AAQ11987.1; -;
DR HSSP; P42312; I89C.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green fl protein.
DR Pfam; PF01353; GFP; 1.

DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 225 AA; 25791 MW; E151D0B497AA23FA CRC64;

Query Match 98.3%; Score 1193; DB 2; Length 225;
Best Local Similarity 98.2%; Pred. No. 2e-96;
Matches 221; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MSCSKNVIKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Qy 61 LSPQFQYGSKVYVVKHPADIPDYKLSFPEGFKWERVMNFEDGGVTVTQDSSLDGCFY 120
Db 61 LSPQFQYGSKVYVVKHPADIPDYKLSFPEGFKWERVMNFEDGGVTVTQDSSLDGCFY 120
Qy 121 KVFIGNVFPSPDGPVMQKTMGWEASTERLYPRDGLVKGEIHKALKLKDGGHYLVEPKSI 180
Db 121 KVFIGNVFPSPDGPVMQKTMGWEASTERLYPRDGLVKGEIHKALKLKDGGHYLVEPKSI 180
Qy 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225
Db 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225

RESULT 4

Q66ND8 Q66ND8 PRELIMINARY; PRT; 236 AA.
AC Q66ND8;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Red fluorescent protein R1.
OS Discosoma sp. RC-2004.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=289055;
RN [1]
RP SEQUENCE FROM N.A.
RA Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT "Cloning of Anthozoan Fluorescent Protein Genes";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY679106; AAU04443.1; -;
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green fl protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 236 AA; 27042 MW; BB3844BCD6834BF3 CRC64;

Query Match 98.1%; Score 1191; DB 2; Length 236;
Best Local Similarity 97.8%; Pred. No. 3.2e-96;
Matches 220; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MSCSKNVIKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Qy 61 LSPQFQYGSKVYVVKHPADIPDYKLSFPEGFKWERVMNFEDGGVTVTQDSSLDGCFY 120
Db 61 LSPQFQYGSKVYVVKHPADIPDYKLSFPEGFKWERVMNFEDGGVTVTQDPSLDGCFY 120
Qy 121 KVFIGNVFPSPDGPVMQKTMGWEASTERLYPRDGLVKGEIHKALKLKDGGHYLVEPKSI 180
Db 121 KVFIGNVFPSPDGPVMQKTMGWEASTERLYPRDGLVKGEIHKALKLKDGGHYLVEPKTI 180
Qy 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225
Db 181 YMAKKPVQLPGYVVYVDSKLDITSHNEDYTIIVEQYTERGHHFL 225

RESULT 5


```
Q9GTJ7
ID Q9GTJ7 PRELIMINARY; PRT; 230 AA.
AC Q9GTJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Red fluorescent protein.
GN Name=RP593;
OS Discosoma sp. SSAL-2000.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=137428;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20434599; PubMed=10981720; DOI=10.1016/S0014-5793(00)01895-0;
RA Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA Lukyanov S.A.;
RT "Novel fluorescent protein from Discosoma coral and its mutants
RT possesses a unique far-red fluorescence.";
RL FEBS Lett. 479:127-130(2000).
DR EMBL; AF272711; AAG16224.1; -.
DR HSSP; Q9U6Y8; 1GGX.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
SQ SEQUENCE 230 AA; 26370 MW; 5215B1B436D67E51 CRC64;

Query Match 89.4%; Score 1085.5; DB 2; Length 230;
Best Local Similarity 88.1%; Pred. No. 5.4e-87;
Matches 199; Conservative 17; Mismatches 9; Indels 1; Gaps 1;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPGYEGHNTVKKLVTKGGPLPFAWDI 60
Db 1 MSCSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPGYEGHNTVKKLVTKGGPLPFAWDI 60

Qy 61 LSPQFQGVSKVYVHPADIPDYKLSPEGFKWERNVNFEDGGVVTVDSSLDGCFIY 120
Db 61 LSPQFQGVSKVYVHPADIPDYKLSPEGFKWERNVNFEDGGVVTVDSSLDGCFIY 120

Qy 121 KVKFIGNVFNPSDGVPMQKTMGWEASTERLYPRDGLVKGTHKALKDGGHYLVFKSI 180
Db 121 EVKFGVNFPSDGVPMQKTMGWEASTERLYPRDGLVKGTHKALKDGGHYLVFKSI 180

Qy 181 YMAKKP-VQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHLFL 225
Db 181 YMVKKPSVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHFFI 226

RESULT 6
Q66PV1 PRELIMINARY; PRT; 221 AA.
ID Q66PV1
AC Q66PV1
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Chromoprotein.
OS Acropora tenuis (Purple tipped acropora).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoenina; Acroporidae; Acropora.
OX NCBI_TaxID=70783;
RN [1]
RP SEQUENCE FROM N.A.
RX Alieva N.O., Melleskevitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646074; AAU06853.1; -.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.

Query Match 65.7%; Score 798; DB 2; Length 221;
Best Local Similarity 69.0%; Pred. No. 8e-62;
Matches 145; Conservative 24; Mismatches 41; Indels 0; Gaps 0;

Qy 6 NVIKEFMRFKVRMEGTVNGHFEIEGEGRPGYEGHNTVKKLVTKGGPLPFAWDILSPQF 65
Db 2 SVIAKQMTYKVMYSGTVNGHYFVEGDKGPKYEGEQTKLVTKGGPLPFAWDILSPQF 61

Qy 66 QYGSKVYVHPADIPDYKLSPEGFKWERNVNFEDGGVVTVDSSLDGCFIYKVKFI 125
Db 62 QYGSIPFTKYDDIPDYKLSPEGFKWERNVNFEDGGVVTVDSSLDGCFIYKVKFI 121

Qy 126 GVNFPDGPVMQKTMGWEASTERLYPRDGLVKGTHKALKDGGHYLVFKSIYMAKK 185
Db 122 GLNFPDGPVMQKTMGWEASTERLYPRDGLVKGTHKALKDGGHYLVFKSIYMAKK 181

Qy 186 PVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHL 223
Db 182 PVRMPGYHYVDRKLDVTNNHNDYTSVEQCEISIAKPL 219

RESULT 7
Q66PUB PRELIMINARY; PRT; 221 AA.
ID Q66PUB
AC Q66PUB
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Chromoprotein.
OS Acropora aculeus.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoenina; Acroporidae; Acropora.
OX NCBI_TaxID=287157;
RN [1]
RP SEQUENCE FROM N.A.
RX Alieva N.O., Melleskevitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646077; AAU06856.1; -.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 24976 MW; 307DE4FC4B018B43 CRC64;

Query Match 65.0%; Score 798; DB 2; Length 221;
Best Local Similarity 69.0%; Pred. No. 8e-62;
Matches 145; Conservative 24; Mismatches 41; Indels 0; Gaps 0;

Qy 6 NVIKEFMRFKVRMEGTVNGHFEIEGEGRPGYEGHNTVKKLVTKGGPLPFAWDILSPQF 65
Db 2 SVIAKQMTYKVMYSGTVNGHYFVEGDKGPKYEGEQTKLVTKGGPLPFAWDILSPQF 61

Qy 66 QYGSKVYVHPADIPDYKLSPEGFKWERNVNFEDGGVVTVDSSLDGCFIYKVKFI 125
Db 62 QYGSIPFTKYDDIPDYKLSPEGFKWERNVNFEDGGVVTVDSSLDGCFIYKVKFI 121

Qy 126 GVNFPDGPVMQKTMGWEASTERLYPRDGLVKGTHKALKDGGHYLVFKSIYMAKK 185
Db 122 GLNFPDGPVMQKTMGWEASTERLYPRDGLVKGTHKALKDGGHYLVFKSIYMAKK 181

Qy 186 PVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHL 223
Db 182 PVRMPGYHYVDRKLDVTNNHNDYTSVEQCEISIAKPL 219

RESULT 8
```

```

Q66PV0
ID Q66PV0 PRELIMINARY; PRT; 221 AA.
AC Q66PV0;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Chromoprotein.
OS Acropora millepora (Coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoeniina; Acroporidae; Acropora.
OX NCBI_TaxID=45264;
RN [1]
RP SEQUENCE FROM N.A.
RA Alieva N.O., Meleshkevitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646075; AAU06854.1; -.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFLUORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
DR SEQUENCE 221 AA; 24987 MW; 546D2CFFAF2552F2 CRC64;

Query Match 65.7%; Score 797; DB 2; Length 221;
Best Local Similarity 68.6%; Pred. No. 9.8e-62;
Matches 144; Conservative 26; Mismatches 40; Indels 0; Gaps 0;

QY 6 NVTKBFMRFKVRMEGTGNGHEFEIEGEGRPRYEGHNTVKLVKTKGGPLPFAMDILSPQF 65
Db SVIAKQMTYKYVMSGTGNGHYFEVQDGKPKYEGEQTVKLVTKGGPLPFAMDILSPQ 61

QY 66 QYGSKYVVKHPADIPDYKKLSPPGFKWERNMFEDGGVVTVDSSLDGCGFIYVKFI 125
Db QYGSIPFTKYPEDIPDYVKQSPFEGYTWERIMNFEDGAVCTVNSDSSIQNCFIYHKFS 121

QY 126 GVNFPSPGPMQKKTGWEASTERLYPRDGLVKGIEHKAALKLDGGHYLVEFKSIYMAKK 185
Db 122 GLNFPNGPMQKKTGWEPTERLFARDGMGLGNFNMALKEGGHYLCEFKSTYKAKK 181

QY 186 PVOLPGYVVDSKLDITSHNEDYIVEQYE 215
Db 182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 211

RESULT 9
Q95P04 PRELIMINARY; PRT; 221 AA.
AC Q95P04;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE GFP-like chromoprotein.
OS Gonopora tenuidens.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Fungiina; Poritidae; Gonopora.
OX NCBI_TaxID=75301;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21538626; PubMed=11682051; DOI=10.1016/S0014-5793(01)02930-1;
RA Gurskaya N.G., Fradkov A.F., Tersikh A., Matz M.V., Labas Y.A.,
RA Martynov V.I., Yanushevich Y.G., Lukyanov K.A., Lukyanov S.A.;
RT "GFP-like chromoproteins as a source of far-red fluorescent
proteins(1).";
RL FEBS Lett. 507:16-20(2001).
DR EMBL; AF383156; AAL27542.1; -.
DR HSP; Q9UY8; 1GGX.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.

```

```

DR PRINTS; PR01229; GFLUORESCENT.
SQ SEQUENCE 221 AA; 24918 MW; 93F9F4B5C2003CB4 CRC64;

Query Match 65.3%; Score 793; DB 2; Length 221;
Best Local Similarity 68.6%; Pred. No. 2.2e-61;
Matches 144; Conservative 25; Mismatches 41; Indels 0; Gaps 0;

QY 6 NVTKBFMRFKVRMEGTGNGHEFEIEGEGRPRYEGHNTVKLVKTKGGPLPFAMDILSPQF 65
Db SVIAKQMTYKYVMSGTGNGHYFEVQDGKPKYEGEQTVKLVTKGGPLPFAMDILSPQ 61

QY 66 QYGSKYVVKHPADIPDYKKLSPPGFKWERNMFEDGGVVTVDSSLDGCGFIYVKFI 125
Db QYGSIPFTKYPEDIPDYVKQSPFEGYTWERIMNFEDGAVCTVNSDSSIQNCFIYHKFS 121

QY 126 GVNFPSPGPMQKKTGWEASTERLYPRDGLVKGIEHKAALKLDGGHYLVEFKSIYMAKK 185
Db 122 GLNFPNGPMQKKTGWEPTERLFARDGMGLGNFNMALKEGGHYLCEFKSTYKAKK 181

QY 186 PVOLPGYVVDSKLDITSHNEDYIVEQYE 215
Db 182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 211

RESULT 10
Q66PU9 PRELIMINARY; PRT; 221 AA.
AC Q66PU9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Chromoprotein.
OS Acropora hvacinthus.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoeniina; Acroporidae; Acropora.
OX NCBI_TaxID=55974;
RN [1]
RP SEQUENCE FROM N.A.
RA Alieva N.O., Meleshkevitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646076; AAU06855.1; -.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR ProDom; PD013756; Green_fl_protein; 1.
DR SEQUENCE 221 AA; 24963 MW; FC9D9E749069D079 CRC64;

Query Match 65.3%; Score 793; DB 2; Length 221;
Best Local Similarity 68.6%; Pred. No. 2.2e-61;
Matches 144; Conservative 25; Mismatches 41; Indels 0; Gaps 0;

QY 6 NVTKBFMRFKVRMEGTGNGHEFEIEGEGRPRYEGHNTVKLVKTKGGPLPFAMDILSPQF 65
Db SVIAKQMTYKYVMSGTGNGHYFEVQDGKPKYEGEQTVKLVTKGGPLPFAMDILSPQ 61

QY 66 QYGSKYVVKHPADIPDYKKLSPPGFKWERNMFEDGGVVTVDSSLDGCGFIYVKFI 125
Db QYGSIPFTKYPEDIPDYVKQSPFEGYTWERIMNFEDGAVCTVNSDSSIQNCFIYHKFS 121

QY 126 GVNFPSPGPMQKKTGWEASTERLYPRDGLVKGIEHKAALKLDGGHYLVEFKSIYMAKK 185
Db 122 GLNFPNGPMQKKTGWEPTERLFARDGMGLGNFNMALKEGGHYLCEFKSTYKAKK 181

QY 186 PVOLPGYVVDSKLDITSHNEDYIVEQYE 215
Db 182 PVKMPGYHYVDRKLDVTNHNKDYTSVEQCE 211

RESULT 11
P83690 PRELIMINARY; PRT; 219 AA.
ID P83690

```


Db 181 YRAKKAALKMPGYHYVDTKLVIWNNDKEFMKVEEHEI A VARHHPF 225

RESULT 14

```

Q963F5 PRELIMINARY; PRT; 225 AA.
ID AC Q963F5
AC Q963F5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastraea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Scleractinia;
OC Faviina; Faviidae; Montastraea.
OX NCBI_TaxID=63558;
[1]
RP SEQUENCE FROM N.A.
RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
RA Falkowski P., Gorbunov M., Kolber Z.;
RL Submitted (AUG-2001) to the ENBL/GenBank/DBJ databases.
DR ENBL; AF384683; AAK62982.2; -.
DR HSSP; Q9U6Y8; IGGX.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP|UOESCENT.
SQ SEQUENCE 225 AA; 25847 MW; 77DE7D7C616929AF CRC64;

```

[illegible]

Search completed: June 30, 2005, 15:49:35
Job time : 176 secs

RESULT 15

Q66ND3	PRELIMINARY	PRT; 225 AA.
Q66ND3		
AC	Q66ND3	
AC	Q66ND3	
DT	25-OCT-2004 (TrEMBLrel. 28, Created)	
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)	
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)	
DE	Green fluorescent protein G2.	
DE	Monstraria cavernosa (great star coral).	
OS	Monstraria cavernosa (great star coral).	
OC	Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia; Scleractinia;	
OC	Favosina; Faviidae; Montastrea.	
OX	NCBI_TaxID=63558;	
OX	[1]	
RN	SEQUENCE FROM N.A.	
RP	Carter R.W., Gibbs P.D.L., Schmale M.C.;	
RA	"Cloning of Cnidarian Fluorescent Protein Genes ";	
RT	Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.	
RL	EMBL; AF679111; AAU04448.1; -.	
DR	InterPro; IPR009017; GPP like.	
DR	InterPro; IPR011584; GPP related.	
DR	InterPro; IPR000786; Green_fl_protein.	
DR	Pfam; PF01353; GPP: 1.	

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:31:23 ; Search time 162 Seconds
(without alignments)
537.167 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSSKNVKEFMFKVRMEG.....EDYIVQEYTERGHHFL 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	100.0	225	3	AY99836 Discosoma
2	1214	100.0	225	3	AB01622 Discosoma
3	1214	100.0	225	4	AG65509 Anthozoan
4	1214	100.0	225	5	AB08834 Yeast opt
5	1214	100.0	225	5	Aae28833 Discosoma
6	1214	100.0	225	5	Aae17540 Discosoma
7	1214	100.0	225	5	Aae18270 Discosoma
8	1214	100.0	225	6	Aae34962 Discosoma
9	1214	100.0	225	7	ADC24126 Discosoma
10	1214	100.0	225	7	ABW00918 Discosoma
11	1214	100.0	225	7	ADF70403 Discosoma
12	1214	100.0	225	7	ADH34489 Discosoma
13	1214	100.0	225	7	ADL46203 Discosoma
14	1214	100.0	225	7	ADN33979 Discosoma
15	1214	100.0	225	8	ADI36421 Discosoma
16	1214	100.0	225	8	ADM97769 Discosoma
17	1214	100.0	225	8	ADQ28780 Jellyfish
18	1214	100.0	487	5	AB08821 Autofluor
19	1214	100.0	506	5	AB08822 Autofluor
20	1214	100.0	547	5	AB08823 Autofluor
21	1211	99.8	225	5	ABE28920 Discosoma
22	1211	99.8	225	7	ABW00937 Discosoma
23	1211	99.8	225	7	ABW00929 Discosoma
24	1211	99.8	225	7	ABW00938 Discosoma
25	1211	99.8	225	7	ADH34498 Discosoma

26	1211	99.8	225	7	ADH34499 Discosoma
27	1210	99.7	225	5	Aae28919 Discosoma
28	1210	99.7	225	7	ABW00930 Discosoma
29	1210	99.7	226	4	AG65510 Anthozoan
30	1210	99.7	242	7	ADZ4109 Discosoma
31	1210	99.7	545	7	ADL18132 RFP:PS(NI
32	1210	99.7	548	7	ADL18156 RFP:PS(HI
33	1209	99.6	225	5	Aae28922 Discosoma
34	1209	99.6	225	7	ABW00936 Discosoma
35	1209	99.6	225	7	ABW00931 Discosoma
36	1209	99.6	226	5	ABB08835 Yeast opt
37	1208	99.5	225	7	ABW00932 Discosoma
38	1208	99.5	225	7	ABW00935 Discosoma
39	1208	99.5	225	7	ABW00939 Discosoma
40	1208	99.5	225	7	ADH34500 Discosoma
41	1207	99.4	225	5	Aae28921 Discosoma
42	1207	99.4	225	5	AAE17541 Discosoma
43	1207	99.4	225	7	ABW00941 Discosoma
44	1207	99.4	225	7	ABW00940 Discosoma
45	1207	99.4	225	7	ADL46222 Discosoma

ALIGNMENTS

RESULT 1

AA99836
ID AA99836 standard; protein; 225 AA.

XX AC AA99836;

XX AC

DT 12-SEP-2003 (revised)

DT 19-SEP-2000 (first entry)

XX Discosoma sp. "red" novel fluorescent protein drFP583.

KW Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism; fluorescent labeling.

XX Discosoma sp; "red".

OS Key Location/Qualifiers

FT Misc-difference 122 /note= "encoded by TC"

FT Misc-difference 127 /note= "encoded by GTTG"

XX WO200034326-A1.

XX 15-JUN-2000.

XX 10-DEC-1999; 99WO-US029473.

XX 11-DEC-1998; 98US-00210330.

XX 14-OCT-1999; 99US-00418529.

XX (CLON-) CLONTECH LAB INC.

XX Lukyanov SA, Fradkov AP, Labas YA, Matz MV, Green G, Chen Y;

XX Ding L;

XX WPI; 2000-423381/36.

XX Novel fluorescent protein from non-bioluminescent Discosoma sp. red, useful for fluorescent labeling and as markers.

XX Claim 20; Page 74-75; 86pp; English.

XX The present sequence is a novel fluorescent protein (nFP) encoded by the full-length cDNA drFP583. drFP583 was isolated from Discosoma sp. "red", a non-bioluminescent species of the Class Anthozoa. Fluorescent proteins can be used in fluorescent labeling, a useful tool for marking a protein, cell or organism of interest. Unlike other markers used in protein

CC labeling, such as beta-galactosidase and luciferase, fluorescent proteins
 CC do not require an exogenous cofactor or substrate. Methods involving
 CC fluorescent proteins are also less laborious and less difficult to
 CC control than the traditional methods of fluorescent labeling, where a
 CC protein of interest is purified and then covalently conjugated to a
 CC fluorophore derivative. Novel fluorescent proteins isolated from species
 CC of the Class Anthozoa can be used as markers for gene expression and
 CC protein localization studies, and in fluorescence resonance energy
 CC transfer (FRET) reactions. They may have improved properties and better
 CC suitability for larger excitations compared to prior art fluorescent
 CC proteins such as green fluorescent protein. (Updated on 12-SEP-2003 to
 CC standardise OS field)

XX SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 3; Length 225;
 Best Local Similarity 100.0%; Pred. No. 8.9e-128;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPYEGHNTVKLVTKGGPLPPAWDI 60
 DB 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPYEGHNTVKLVTKGGPLPPAWDI 60
 QY 61 LSPQFOYGSKVYVHPADIPDYKLSFPEGFKWERVMNFDGGVVTVTQDSSLQDGCIFY 120
 DB 61 LSPQFOYGSKVYVHPADIPDYKLSFPEGFKWERVMNFDGGVVTVTQDSSLQDGCIFY 120
 QY 121 KVKEFGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGGHYLVPEFSI 180
 DB 121 KVKEFGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGGHYLVPEFSI 180
 QY 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
 DB 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 2

AAB01622
 ID AAB01622 standard; protein; 225 AA.

XX AC AAB01622;

XX DT 12-DEC-2000 (first entry)

XX DE Diacosoma sp. red fluorescent protein drFP583.

XX KW Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;
 XX KW zFP506; zFP538; drFP583; drFP483; drFP600; drFP512; drFP592.

XX OS Diacosoma sp.

XX PN WO200034526-A1.

XX PD 15-JUN-2000.

XX PF 10-DEC-1999; 99WO-US029405.

XX PR 11-DEC-1998; 98US-00210330.

XX PA (CLON-) CLONTECH LAB INC.

XX PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV;

XX DR WPI; 2000-423451/36.

XX PT Novel method for identifying a DNA sequence encoding fluorescent proteins
 XX PT from non-bioluminescent Anthozoa which are useful for fluorescent
 XX PT labeling and as markers.

XX PS Claim 3; Page 68-69; 73pp; English.

XX CC The present sequence is Diacosoma sp. red fluorescent protein drFP583. It
 CC was isolated using the Aequoria victoria green fluorescent protein (GFP)

CC sequence, which was used to design PCR primers which might isolate other
 CC fluorescent proteins from a number of species of Anthozoa. These were
 CC Anemonia majano, Clavularia sp., Zoanthus sp., Diacosoma sp. and Anemonia
 CC sulcata. The cDNA obtained was then screened in the search for sequences
 CC encoding fluorescent proteins. The other proteins found in this manner
 CC were cFP484, zFP506, zFP538, amFP486, drFP600, drFP512 and
 CC drFP592. These proteins can be used as fluorescent labels (for gene
 CC expression and protein localisation studies and in fluorescence resonance
 CC energy transfer (FRET) studies) in place of fluorophore derivatives and
 CC luciferases, as these involve laborious processes and the latter require
 CC cofactors. They can also be used in place of GFP, which is too stable to
 CC be useful when studying short-term or repetitive events

XX SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 3; Length 225;
 Best Local Similarity 100.0%; Pred. No. 8.9e-128;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPYEGHNTVKLVTKGGPLPPAWDI 60
 DB 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPYEGHNTVKLVTKGGPLPPAWDI 60
 QY 61 LSPQFOYGSKVYVHPADIPDYKLSFPEGFKWERVMNFDGGVVTVTQDSSLQDGCIFY 120
 DB 61 LSPQFOYGSKVYVHPADIPDYKLSFPEGFKWERVMNFDGGVVTVTQDSSLQDGCIFY 120
 QY 121 KVKEFGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGGHYLVPEFSI 180
 DB 121 KVKEFGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGGHYLVPEFSI 180
 QY 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
 DB 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 3

AAG65509
 ID AAG65509 standard; protein; 225 AA.

XX AC AAG65509;

XX DT 30-NOV-2001 (first entry)

XX DE Anthozoan red fluorescent protein sequence.

XX KW Fluorescent protein; Anthozoan; fluorescence; marker; FRET; red.

XX OS Anthozoa.

XX PN WO200162919-A1.

XX PD 30-AUG-2001.

XX PF 13-FEB-2001; 2001WO-US0004625.

XX PR 23-FEB-2000; 2000US-0184732P.

XX PA (AURO-) AURORA BIOSCIENCES CORP.

XX PI Nelson D, Zamaira E, Tsien R;

XX DR WPI; 2001-557704/62.

XX PT Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise
 XX PT functional red fluorescent proteins, and the encoding nucleic acids, with
 XX PT key mutations for improving the proteins function.

XX PS Disclosure; Page 85; 90pp; English.

XX CC The invention provides a nucleic acid encoding functional red fluorescent
 CC protein (II) that differs from the sequence of an Anthozoan red
 CC fluorescent protein by at least one amino acid substitution, and with

CC different fluorescent properties. The red fluorescent protein of the
CC invention can be expressed by standard recombinant methodology. (II) are
CC used a fluorescent markers and FRET partners. It is used for identifying
CC protein-protein interactions. (II) is also suitable for multiplexed
CC fluorescent analysis and FRET-based applications using existing Aequorea
CC fluorescent proteins. (II) has improved brightness, reduced spectral
CC cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC The key mutations in the encoding nucleic acids provide improved folding,
CC brightness, and create (II) with sharper, more defined excitation and
CC emission peaks when expressed in mammalian cells. The present sequence
CC represents an anthozoan fluorescent protein
XX
SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 4; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRSSKNVKEFMFKVMECTVNGHEFEIEGEGRPEYEGHNTVKLVKVTGGPLPFAWDI 60
DB 1 MRSSKNVKEFMFKVMECTVNGHEFEIEGEGRPEYEGHNTVKLVKVTGGPLPFAWDI 60
QY 61 LSPQFQYGSKVYVKGHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
DB 61 LSPQFQYGSKVYVKGHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
QY 121 KVPFIGNPSPDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGSHYLVEPKSI 180
DB 121 KVPFIGNPSPDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGSHYLVEPKSI 180
QY 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHLFL 225
DB 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHLFL 225

RESULT 4
AB08834
ID ABB08834 standard; protein; 225 AA.
XX AC ABB08834;
XX DT 29-MAY-2002 (first entry)
XX DE Yeast optimised RFP SEQ ID NO 17.
XX KW Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
XX KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
XX KW Escherichia coli; green fluorescent protein; biotechnology.
XX OS Anthozoa.
XX PN DE20001395-U1.
XX PD 15-MAR-2001.
XX PF 27-JAN-2000; 2000DE-02001395.
XX PR 27-JAN-2000; 2000DE-02001395.
XX PA (GPCB-) GPC BIOTECH AG.
XX DR WPI; 2002-228394/29.
XX DR N-PSDB; ABA95905, ABA95921, ABA95922.
XX PT New DNA encoding red fluorescent protein, useful as marker in
XX PT biotechnology, has sequence optimized for expression in eukaryotes,
XX PT especially yeast or plants.
XX PS Disclosure; Page 13-14; 19pp; German.
XX CC The invention relates to DNA (I) containing either sequence ABA95905 or
XX CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
XX CC (YRFP). (I) are used to express red fluorescent protein (RFP) in

CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC plants, especially dicotyledonous plants including Nicotiana tabacum or
CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC especially Escherichia coli. RFP is useful in the same way as green
CC fluorescent protein but is more generally applicable in modern
CC biotechnology. (I) are optimised for expression in yeast and so generate
CC RFP at higher levels with stronger fluorescence and thus lowers the
CC detection limit and gives a better signal-to-noise ratio. The present
CC sequence is that of the yeast optimised RFP
XX
SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRSSKNVKEFMFKVMECTVNGHEFEIEGEGRPEYEGHNTVKLVKVTGGPLPFAWDI 60
DB 1 MRSSKNVKEFMFKVMECTVNGHEFEIEGEGRPEYEGHNTVKLVKVTGGPLPFAWDI 60
QY 61 LSPQFQYGSKVYVKGHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
DB 61 LSPQFQYGSKVYVKGHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
QY 121 KVPFIGNPSPDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGSHYLVEPKSI 180
DB 121 KVPFIGNPSPDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLDGSHYLVEPKSI 180
QY 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHLFL 225
DB 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYTERGHRHLFL 225

RESULT 5
AAE28833
ID AAE28833 standard; protein; 225 AA.
XX AC AAE28833;
XX DT 27-DEC-2002 (first entry)
XX DE Discosoma sp. drFP583 (NFP-6) wild-type protein.
XX KW Fluorescent protein; chromoprotein; protease cleavage assay; filter;
XX KW fluorescence activated cell sorting application; fluorescent timer;
XX KW biosensor; fluorescence resonance energy transfer application; FRET;
XX KW colouring agent; recombinant DNA application; analyte detection assay;
XX KW sunscreen; second messenger detector; drFP583 protein; NFP-6.
XX OS Discosoma sp.
XX PN WO200268459-A2.
XX PD 06-SEP-2002.
XX PF 20-FEB-2002; 2002WO-US005749.
XX PR 21-FEB-2001; 2001US-0270983P.
XX PR 04-DEC-2001; 2001US-00006922.
XX PA (CLON-) CLONTECH LAB INC.
XX PI Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX DR WPI; 2002-691654/74.
XX DR N-PSDB; AAD46278.
XX PT New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX PT of an aggregating Chidarian chromo- or fluorescent protein or mutant for
XX PT analyte detection assays or fluorescence activated cell sorting
XX PT applications.
XX PS Disclosure; Page 70-71; 80pp; English.

XX The invention relates to nucleic acid molecules encoding non-aggregating
CC chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
CC useful in analyte detection assays, as colouring agents, as markers in
CC recombinant DNA applications, as sunscreens or filters, in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, in screening assays, as second
CC messenger detectors, in fluorescence activated cell sorting applications,
CC in protease cleavage assays or as fluorescent timers. The present
CC invention is *Discosoma* sp. drFP583 (NFP-6) wild-type protein of the
XX sequence 225 AA;
SQ Sequence 225 AA;
Query Match 100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRSSKNVKEFMRFKVRMEGTGNGHEFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTGNGHEFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Qy 61 LSPQFQYGSKVYVYKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLDGCFYI 120
Db 61 LSPQFQYGSKVYVYKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLDGCFYI 120
Qy 121 KVKFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLKDGGHVLYVEFKSI 180
Db 121 KVKFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLKDGGHVLYVEFKSI 180
Qy 181 YMAKKPVQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
RESULT 6
AAE17540
ID AAE17540 standard; protein; 225 AA.
XX AAE17540;
AC AAE17540;
XX 22-APR-2002 (first entry)
XX *Discosoma* sp. humanised wild-type Anthozoa protein drFP583.
XX Fluorescent timer protein; protein movement; translocation; trafficking;
KW promoter activity; gene expression; transgenic plant; gene modification;
KW protein age; anthozoa protein; drFP583.
XX *Discosoma* sp.
XX W0200196373-A2.
XX 20-DEC-2001.
XX 13-JUN-2001; 2001WO-US019097.
XX 14-JUN-2000; 2000US-0211607P.
XX (CLON-) CLONTECH LAB INC.
XX Fradkov AF, Tersikh A;
XX WPI; 2002-154595/20.
DR N-PSDB; AAD28207.
XX New fluorescent timer proteins comprising an emission spectrum that
PT changes over time from a first wavelength to a second wavelength, useful
PT for monitoring intracellular protein movement, translocation, trafficking
PT or stability.
XX Example 1; Fig 1; 89pp; English.

CC The invention relates to a fluorescent timer protein having an emission
CC spectrum that changes over time after synthesis from a first wavelength
CC to a second wavelength. The fluorescent timer proteins are useful in
CC monitoring the activity of a promoter, determining the age of a protein,
CC identifying an agent that modulates the activity of a promoter and in
CC enriching a population of cells comprising a fluorescent timer protein.
CC The fluorescent timer proteins are also useful for assessing gene
CC expression during development of a multicellular organism or during
CC cellular differentiation, in response to a drug or other inducer of
CC promoter activity, as a reporter to serve as a read-out of promoter
CC activity, monitoring intracellular protein movement or translocation,
CC protein trafficking, or protein stability, to investigate temporal
CC aspects of the activity of a regulatory element, for determining cell
CC fate during development and organ remodelling, in spatial and temporal
CC visualisation of newly synthesised proteins and accumulated proteins, and
CC in distinguishing between newly formed and pre-existing structures, e.g.
CC membrane junctions and extracellular matrix components. The fluorescent
CC timer proteins may further be used to investigations where photobleaching
CC techniques are employed, as detectable labels, as selectable markers, as
CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
CC assays, and as second messenger detectors. The nucleic acids can be used
CC to generate transgenic, non-human plants or animals or site-specific gene
CC modifications in cell lines. The present sequence is *Discosoma* sp.
CC humanised wild-type Anthozoa protein drFP583 used for generating
XX fluorescent proteins
SQ Sequence 225 AA;
Query Match 100.0%; Score 1214; DB 5; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRSSKNVKEFMRFKVRMEGTGNGHEFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTGNGHEFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Qy 61 LSPQFQYGSKVYVYKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLDGCFYI 120
Db 61 LSPQFQYGSKVYVYKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLDGCFYI 120
Qy 121 KVKFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLKDGGHVLYVEFKSI 180
Db 121 KVKFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHEIKALKLKDGGHVLYVEFKSI 180
Qy 181 YMAKKPVQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
RESULT 7
AAO18270
ID AAO18270 standard; protein; 225 AA.
XX AAO18270;
AC AAO18270;
XX 26-SEP-2002 (first entry)
XX *Discosoma* red fluorescent protein.
XX Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
KW modified yeast strain; environmental pollution.
XX *Discosoma* sp.
XX DE10061872-A1.
XX 20-JUN-2002.
XX 12-DEC-2000; 2000DE-01061872.
XX 12-DEC-2000; 2000DE-01061872.
XX (LICH/) LICHTENBERG-FRATE H.
XX

XX PI Lichtenberg-Frate H;
 XX DR WPI; 2002-539633/58.
 XX DR N-PSDB; AAL47952.
 XX PT Modified yeast strain, useful for detecting toxic compounds in
 PT environment, contains integrated cassettes responsive to genotoxic and
 PT cytotoxic compounds.
 XX PS Disclosure; Page 21-22; 34pp; German.
 XX CC The present invention relates to a modified yeast strain that contains,
 CC integrated stably and functionally in its genome, a genotoxicity cassette
 CC and a cytotoxicity cassette, each comprising a promoter and reporter
 CC gene, both of which are different in the two cassettes. The modified
 CC yeast strain is used to detect environmental pollution, especially
 CC genotoxic and/or cytotoxic substances in complex environmental
 CC contaminants, especially organic compounds, but also (non-)ionising
 CC radiation and chemical carcinogens. Particular applications are in
 CC monitoring (waste) water (e.g. as an early warning system), medical
 CC toxicology screening and for industrial process control. The present
 CC sequence is a marker protein suitable for use in the cassettes of the
 CC present invention
 XX SQ Sequence 225 AA;
 Query Match 100.0%; Score 1214; DB 5; Length 225;
 Best Local Similarity 100.0%; Pred. No. 8.9e-128;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRSSKNVKEFMFVKRMETVNGHEPEIEGEGRPEYEGHNTVKLVKGGPLPFAWDI 60
 Db 1 MRSSKNVKEFMFVKRMETVNGHEPEIEGEGRPEYEGHNTVKLVKGGPLPFAWDI 60
 QY 61 LSPQFQYGSVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQDSSLQDGCIFY 120
 Db 61 LSPQFQYGSVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQDSSLQDGCIFY 120
 QY 121 KVXFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHIHAKLKDGGHYLVEPKSI 180
 Db 121 KVXFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHIHAKLKDGGHYLVEPKSI 180
 QY 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTYVEQVTEGRHHFL 225
 Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTYVEQVTEGRHHFL 225
 RESULT 8
 AAE34962
 ID AAE34962 standard; protein; 225 AA.
 XX AC AAE34962;
 XX DT 28-MAY-2003 (first entry)
 XX DE Discosoma species red fluorescent protein (RFP).
 XX KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW kinase; red fluorescent protein; RFP.
 XX OS Discosoma sp.
 XX PN WO200295058-A2.
 XX PD 28-NOV-2002.
 XX PF 24-MAY-2002; 2002WO-US016955.
 XX PR 24-MAY-2001; 2001US-00865291.
 XX PA (REGC) UNIV CALIFORNIA.
 XX

PI Tsien RY, Ting AY, Zhang J;
 DR WPI; 2003-148474/14.
 DR N-PSDB; AADS3432.
 XX PT Novel chimeric phosphorylation indicators, useful for detecting
 PT kinase/phosphatase in samples, has donor molecule, phosphorylatable
 PT domain, phosphaminoacid binding domain, and acceptor molecule, in
 PT operative linkage.
 XX PS Disclosure; Col 65-66; 38pp; English.
 XX CC The present invention relates to chimeric phosphorylation indicators
 CC comprising a phosphorylation polypeptide and a fluorescent protein or in
 CC operative linkage, a donor molecule, a phosphorylatable domain, a
 CC phosphaminoacid binding domain (PABD) and an acceptor molecule. The
 CC phosphorylation indicators of the invention are useful for detecting
 CC kinases or phosphatases in a biological sample. They are also useful in
 CC high throughput analysis e.g. for detecting a kinase inhibitor or
 CC phosphatase inhibitor. The present sequence is Discosoma species red
 CC fluorescent protein (RFP) used in the invention
 XX SQ Sequence 225 AA;
 Query Match 100.0%; Score 1214; DB 6; Length 225;
 Best Local Similarity 100.0%; Pred. No. 8.9e-128;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MRSSKNVKEFMFVKRMETVNGHEPEIEGEGRPEYEGHNTVKLVKGGPLPFAWDI 60
 Db 1 MRSSKNVKEFMFVKRMETVNGHEPEIEGEGRPEYEGHNTVKLVKGGPLPFAWDI 60
 QY 61 LSPQFQYGSVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQDSSLQDGCIFY 120
 Db 61 LSPQFQYGSVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQDSSLQDGCIFY 120
 QY 121 KVXFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHIHAKLKDGGHYLVEPKSI 180
 Db 121 KVXFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGHIHAKLKDGGHYLVEPKSI 180
 QY 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTYVEQVTEGRHHFL 225
 Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTYVEQVTEGRHHFL 225
 RESULT 9
 ADC24126
 ID ADC24126 standard; protein; 225 AA.
 XX AC ADC24126;
 XX DT 18-DEC-2003 (first entry)
 XX DE Discosoma wild-type red fluorescent protein.
 XX KW Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
 KW fluorescent protein variant; transcription induction detection;
 KW fluorescence energy resonance transfer; FRET; protein kinase;
 KW protein phosphatase; ion indicator.
 XX OS Discosoma.
 XX PN US2003059835-A1.
 XX PD 27-MAR-2003.
 XX PF 10-APR-2002; 2002US-00121258.
 XX PR 26-FEB-2001; 2001US-00794308.
 XX PR 24-MAY-2001; 2001US-00866538.
 XX PA (TSIE/) TSIE R Y.
 PA (CAMP/) CAMPBELL R E.

```
XX Tsien RY, Campbell RE;
PI
XX
XX
DR WPI; 2003-743764/70.
DR N-PSDB; ADC241127, ADC241134.
XX
XX Novel polynucleotide sequence encoding Discosoma red fluorescent protein
PT variant having a reduced propensity to oligomerize, useful for detecting
PT transcriptional activity.
XX
XX Claim 1; SEQ ID NO 1; 67pp; English.
XX
XX The invention describes a polynucleotide sequence (I) encoding a
CC Discosoma red fluorescent protein (DsRed) variant having a reduced
CC propensity to oligomerize, comprising amino acid substitutions at the AB
CC and/or AC interfaces of the wild-type DsRed sequence (SI) comprising 225
CC amino acids, given in the specification, where the substitutions result
CC in reduced propensity of the DsRed variant to form tetramers. (I) is
CC useful for detecting transcriptional activity by providing a host cells
CC containing a vector which comprises (I) operatively linked to an
CC expression control sequence, and an unit to assay the variant fluorescent
CC protein fluorescence, and assaying fluorescence of the variant
CC fluorescent protein produced by (VII), where variant fluorescent protein
CC fluorescence is indicative of transcriptional activity. A polynucleotide
CC encoding a fusion protein is useful for the analysis of in vivo
CC localisation or trafficking of a polypeptide of interest. A polypeptide
CC marker is useful as markers to identify the location and amount of a
CC target protein produced, where the target protein is fused to the marker,
CC as a complement to or alternative for the green fluorescent protein or
CC its spectral variant, for detecting induction of transcriptions, in
CC applications involving fluorescence energy resonance transfer (FRET),
CC which detects events as the function of the movement of fluorescent
CC donors and acceptors towards or away from each other, for making
CC fluorescent sensors for protein kinase and phosphatase activities or
CC indicators for ions and molecules such as  $Ca^{2+}$ ,  $Zn^{2+}$ , for identifying the
CC presence of a molecule in a sample, for identifying a specific
CC interaction of a first and second molecule, for determining whether a
CC sample contains an enzyme or for determining the pH of the sample. (I) is
CC useful for identifying a region or condition that regulates the activity
CC of an expression control sequence. This is the amino acid sequence of
CC Discosoma wild-type red fluorescent protein.
XX
SQ Sequence 225 AA;
Query Match 100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRSSKNVKEFMRFKVRMEGTVNGHGFEPFIEGEGEPYEGHNTVKLVKTKGGPLPFAWDI 60
DB 1 MRSSKNVKEFMRFKVRMEGTVNGHGFEPFIEGEGEPYEGHNTVKLVKTKGGPLPFAWDI 60
QY 61 LSPQFQYGSKYVVKHPADIPDYKLSPPGFKWERNVNFEDGGVTVTQDSSLDQGCIFY 120
DB 61 LSPQFQYGSKYVVKHPADIPDYKLSPPGFKWERNVNFEDGGVTVTQDSSLDQGCIFY 120
QY 121 KVKFIGNVFPDGPVNMOKKTGMWEASTERLYPRDGLVKGKGIHKAALKDKGGHYLVFPKSI 180
DB 121 KVKFIGNVFPDGPVNMOKKTGMWEASTERLYPRDGLVKGKGIHKAALKDKGGHYLVFPKSI 180
QY 181 YMAKKPVQLPGYYVDSKLDITSHNEDYTIQEYTERGRRHLFL 225
DB 181 YMAKKPVQLPGYYVDSKLDITSHNEDYTIQEYTERGRRHLFL 225
RESULT 11
ADF70403
ID ADF70403 standard; protein; 225 AA.
XX
XX ADF70403;
AC
XX
XX 12-FEB-2004 (first entry)
DT
XX
XX Discosoma wild-type GFP variant protein SeqID26.
DE
XX
XX ligand; orphan receptor protein; fusion protein; fluorescent protein;
```

KW cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
KW GFPuv; Enhanced GFP; EGFP.
XX Discosoma sp.
XX WO2003071272-A1.
XX
XX 28-AUG-2003.
XX
XX 21-FEB-2003; 2003WO-JP001901.
XX
XX 22-FEB-2002; 2002JP-00045728.
XX 23-JUL-2002; 2002JP-00213949.
PR 11-OCT-2002; 2002JP-00298237.
XX
XX (TAKE) TAKEDA CHEM IND LTD.
XX
XX Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX WPI; 2003-697654/66.
DR N-PSDB; ADF70404.
XX
XX Transformation of cells with a fusion protein of an orphan receptor
PT protein with a fluorescent protein useful for identification of ligands
PT to the orphan receptor.
XX
XX Disclosure; SEQ ID NO 26; 594pp; Japanese.
XX
XX This invention relates to a novel method of identifying ligands to an
CC orphan receptor protein which comprises transforming cells with DNA
CC encoding a fusion protein of the orphan receptor with a fluorescent
CC protein, so that the fusion protein is expressed in the cells (or cell
CC membranes isolated from them) and contacting the cells with the potential
CC ligand to be tested. A suitable fluorescent protein for incorporation in
CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,
CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
CC identification of ligands binding to an orphan receptor protein.
XX
XX Sequence 225 AA;
SQ
Query Match 100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRSKNVKEFMFKVMECTVNGHEFEIEGEGEPYEGHNTVKLVKGGPLPPAWDI 60
DB 1 MRSKNVKEFMFKVMECTVNGHEFEIEGEGEPYEGHNTVKLVKGGPLPPAWDI 60
QY 61 LSPQFYGSKVYKHPADIPDYKLSPEGFKWERNVFNEDGGVVTVTQDSSLDQGCIFY 120
DB 61 LSPQFYGSKVYKHPADIPDYKLSPEGFKWERNVFNEDGGVVTVTQDSSLDQGCIFY 120
QY 121 KVXFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGIHKAALKDKGGHYLVPEKSI 180
DB 121 KVXFIGVNFPSDGPVMQKTMGWEASTERLYPRDGLVKGIHKAALKDKGGHYLVPEKSI 180
QY 181 YMAKPKVQLPGYYVDSKLDITSHNEDYTYVEQYTERGHHFL 225
DB 181 YMAKPKVQLPGYYVDSKLDITSHNEDYTYVEQYTERGHHFL 225
RESULT 12
ADH34489
ID ADH34489 standard; protein; 225 AA.
XX
XX ADH34489;
AC
XX
DT 11-MAR-2004 (first entry)
XX
XX Discosoma sp. red fluorescent protein DsRed (wild-type).
XX Chromoprotein; fluorescent protein; CP; FP; interconverted mutant;
KW Chidnarian; Anthozoan; labelling; colouring agents; pigment;

KW analyte detection assay; selectable marker; sunscreen; selective filter;
KW fluorescence resonance energy transfer; FRET; biosensor;
KW whole cell marker; second messenger detector; in vivo marker;
KW fluorescence activated cell sorting; fluorescent timer;
KW red fluorescent protein; DsRed.
XX
XX Discosoma sp.
XX
XX Key Location/Qualifiers
FH Misc-difference 146
FT /note= "Key position for chromoprotein/fluorescent
FT protein activity. Corresponds to GFP residue 148"
FT Misc-difference 161
FT /note= "Key position for chromoprotein/fluorescent
FT protein activity. Corresponds to GFP residue 165"
FT Misc-difference 163
FT /note= "Key position for chromoprotein/fluorescent
FT protein activity. Corresponds to GFP residue 167"
FT Misc-difference 197
FT /note= "Key position for chromoprotein/fluorescent
FT protein activity. Corresponds to GFP residue 203"
XX
XX WO2003057833-A2.
XX
XX 17-JUL-2003.
XX
XX 23-DEC-2002; 2002WO-US041418.
XX
XX 26-DEC-2001; 2001US-0343128P.
XX (CLON-) CLONTECH LAB INC.
XX
XX Bulina MB, Chudakov D, Lukyanov KA;
XX WPI; 2003-607998/57.
XX
XX Novel nucleic acid encoding interconverted mutant of chromo-or
PT fluorescent protein which are useful as biosensors, coloring agents.
XX
XX Example 1; Fig 1; 56pp; English.
XX
XX The invention relates to interconverted mutants of chromoproteins (CP) or
CC fluorescent proteins (FP) and nucleic acids encoding them. The mutant is
CC derived from a Chidnarian species, preferably a non-bioluminescent
CC Chidnarian species, and most preferably an Anthozoan species. The
CC invention is based on the finding that although green fluorescent protein
CC (GFP)-like chromoproteins and fluorescent proteins exhibit some degree of
CC homology, there are certain positions (referred to as 148, 165, 167 and
CC 203; numbering corresponds to GFP) that are occupied by noticeably
CC different residues in the two types of proteins. Mutagenesis of the
CC residues in these key positions in, for example, a fluorescent protein,
CC to those found in a chromoprotein is therefore proposed to confer
CC chromoprotein activity on the fluorescent protein mutant, with
CC chromoproteins being able to be converted into fluorescent proteins in a
CC similar manner. The invention also relates to expression constructs,
CC vectors, host cells and host cell progeny comprising a nucleic acid of
CC the invention; the recombinant production of an interconverted
CC chromoprotein or fluorescent protein mutant; and antibodies specific for
CC interconverted mutant proteins of the invention. The interconverted
CC mutants are useful in any application that employs a chromoprotein or
CC fluorescent protein. Fluorescent protein mutants having chromoprotein
CC activity can be useful as coloring agents in, for example, food
CC compositions, pharmaceuticals, cosmetics and living organisms. Proteins
CC with chromoprotein activity are also useful as labels in biological
CC applications (e.g. the production of transgenic cells and organisms), and
CC analyte detection assays, as selectable markers in recombinant DNA
CC are also useful as sunscreens and selective filters. Chromoprotein
CC mutants having fluorescent protein activity useful in fluorescence
CC resonance energy transfer (FRET) applications, as biosensors in
CC prokaryotic and eukaryotic cells, as markers of whole cells to detect
CC changes in multicellular reorganization and migration, as second
CC messenger detectors, as in vivo markers in animals (e.g., transgenic
CC animals), in fluorescence activated cell sorting applications, in

CC protease cleavage assays, and in assays to determine the phospholipid
CC composition in biological membranes. Proteins with fluorescent protein
CC activity can also be used as fluorescent timers, where the switch of one
CC fluorescent colour to another (e.g., green to red) is concomitant with
CC the ageing of the protein and is useful for determination of the
CC activation or deactivation of gene expression. The present sequence
CC represents a wild-type red fluorescent protein, DsRed, from *Discosoma* sp.
CC that was used as a parent sequence for the generation of mutant proteins
CC in an example of the invention.

XX
SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
DB 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
QY 61 LSPQFOYGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFFIY 120
DB 61 LSPQFOYGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFFIY 120
QY 121 KVKFIGVNFPSDGPVQKKTGMWEASTERLYPRDGLKGEIHKALKLKDGGHYLVFVKSI 180
DB 121 KVKFIGVNFPSDGPVQKKTGMWEASTERLYPRDGLKGEIHKALKLKDGGHYLVFVKSI 180
QY 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIQEYTERGRRHHLFL 225
DB 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIQEYTERGRRHHLFL 225

RESULT 13

ADL46203
ID ADL46203 standard; protein; 225 AA.

XX AC ADL46203;

XX AC ADL46203;

DT 20-MAY-2004 (first entry)

XX *Discosoma* red fluorescent (DsRed) protein.

DE red fluorescent protein; DsRed; fluorescence; red wavelength;
XX oligomerization; tetramerization; immunoassay; hybridization assay.

XX *Discosoma* sp.

XX WO2003086446-A1.

XX 23-OCT-2003.

XX 09-APR-2003; 2003WO-US010879.

XX 10-APR-2002; 2002US-00121258.

XX 29-JUL-2002; 2002US-00209208.

XX (REGC) UNIV CALIFORNIA.

XX Tsien RY, Campbell RE, Baird GS;

XX WPI; 2003-845265/78.

XX N-PSDB; ADL46204.

XX New monomeric and dimeric Anthozoan fluorescent protein variants with
PT reduced propensity to oligomerize, and encoding polynucleotides, useful
PT in molecular biology, e.g. in immunoassays or in tracking protein
PT movement in cells.

XX Claim 1; SEQ ID NO 1; 166pp; English.

XX The invention relates to a polynucleotide sequence encoding a *Discosoma*
CC red fluorescent protein (DsRed) variant having a reduced propensity to

CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DsRed
CC sequence, where the substitutions result in reduced propensity of the
CC DsRed variant to form tetramers and where the variant displays detectable
CC fluorescence of at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the DsRed
CC protein.

XX
SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 7; Length 225;
Best Local Similarity 100.0%; Pred. No. 8.9e-128;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
DB 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
QY 61 LSPQFOYGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFFIY 120
DB 61 LSPQFOYGSKVYVKKHPADIPDYKKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCFFIY 120
QY 121 KVKFIGVNFPSDGPVQKKTGMWEASTERLYPRDGLKGEIHKALKLKDGGHYLVFVKSI 180
DB 121 KVKFIGVNFPSDGPVQKKTGMWEASTERLYPRDGLKGEIHKALKLKDGGHYLVFVKSI 180
QY 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIQEYTERGRRHHLFL 225
DB 181 YMAKPVQLPGYVYVDSKLDITSHNEDYTIQEYTERGRRHHLFL 225

RESULT 14

ADN33979
ID ADN33979 standard; protein; 225 AA.

XX AC ADN33979;

DT 01-JUL-2004 (first entry)

XX Wild-type DsRed protein.

DE Chnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRed.

XX *Discosoma* sp.

XX WO2003054158-A2.

XX 03-JUL-2003.

XX 18-DEC-2002; 2002WO-US040539.

XX 19-DEC-2001; 2001US-0341723P.

XX (UYCH-) UNIV CHICAGO.

XX Bevis B, Glick B;

XX WPI; 2003-569236/53.

XX N-PSDB; ADN33978.

XX Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT mutant of a Chnidarian chromo- or fluorescent protein or its mutant,
PT useful for applications involving chromo- or fluorescent proteins.

XX Claim 8; SEQ ID NO 2; 65pp; English.

XX The present invention relates to nucleic acid that encodes a rapidly
CC maturing chromo or fluorescent mutant of a Chnidarian chromo- or
CC fluorescent protein or its mutant. The protein is useful in applications

CC involving nucleic acid encoding a chromo- or fluorescent protein and is
 CC useful for producing a chromo and/or fluorescent protein which involves
 CC growing the cell, whereby the protein is expressed, and isolating the
 CC protein substantially free of other proteins. The protein is useful in
 CC applications involving chromo- or fluorescent protein and is useful as
 CC PCR primers, hybridization probes, etc. The expression cassettes are
 CC useful for synthesizing related proteins. The chromoproteins are useful
 CC as coloring agents which are capable of imparting color or pigment to a
 CC particular composition of matter e.g. food compositions, pharmaceuticals,
 CC cosmetics, living organisms, e.g., animals and plants. The chromoproteins
 CC may also find use as labels in analyte detection assays, e.g. assays for
 CC biological analytes of interest and as selectable markers in recombinant
 CC DNA applications, e.g. the production of transgenic cells and organisms.
 CC The fluorescent proteins find use in a variety of different applications,
 CC e.g. in fluorescence resonance energy transfer (FRET) applications, as
 CC biosensors in prokaryotic and eukaryotic cells, in applications involving
 CC the automated screening of arrays of cells expressing fluorescent
 CC reporting groups by using microscopic imaging and electronic analysis, as
 CC second messenger detectors, and in fluorescence activated cell sorting
 CC applications and as in vivo marker in animals. The fluorescent proteins
 CC also find use in protease cleavage assays. The proteins can also be used
 CC is assays to determine the phospholipid composition in biological
 CC membranes and as a fluorescent timer. The present sequence represents the
 CC wild-type DsRED.

XX Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 7; Length 225;
 Best Local Similarity 100.0%; Pred. No. 8.9e-128;
 Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMECTVNGHEFEIEGEGRPEYGHNTVKLVKTKGGPLPFAWDI 60
 DB 1 MRSSKNVKEFMRFKVRMECTVNGHEFEIEGEGRPEYGHNTVKLVKTKGGPLPFAWDI 60

QY 61 LSPQFOYGVSKVYVKKHPADIPDYKKLSPEGFKWERNVNFEDGGVVTQTSSLDQGCIFY 120
 DB 61 LSPQFOYGVSKVYVKKHPADIPDYKKLSPEGFKWERNVNFEDGGVVTQTSSLDQGCIFY 120

QY 121 KVKFIGVNFPSDGPVNMQKTMGWEASTERLYPRDGLVKGEIHKALKDKGGHYLVPEFKSI 180
 DB 121 KVKFIGVNFPSDGPVNMQKTMGWEASTERLYPRDGLVKGEIHKALKDKGGHYLVPEFKSI 180

QY 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTYVEQYTERGHRHLFL 225
 DB 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTYVEQYTERGHRHLFL 225

RESULT 15

ADI36421
 ID ADI36421 standard; protein; 225 AA.

XX AC ADI36421;

XX DT 22-APR-2004 (first entry)

XX DE Discosoma sp. red fluorescent protein (RED).

XX KW Fluorobody; binding ligand; green fluorescent protein; GFP;
 XX target detection; red fluorescent protein; RED.

XX OS Discosoma sp.

XX Key Location/Qualifiers
 XX Region 22..26
 XX Region /note= "Loop region"
 XX Region 37..39
 XX Region /note= "Loop region"
 XX Region 75..81
 XX Region /note= "Loop region"
 XX Region 100..103
 XX Region /note= "Loop region"
 XX Region 114..117

FT Region /note= "Loop region"
 FT 153..156
 FT Region /note= "Loop region"
 FT 167..170
 FT Region /note= "Loop region"
 FT 185..192
 FT Region /note= "Loop region"
 FT 204..209
 FT Region /note= "Loop region"

XX US2003203355-A1.

XX 30-OCT-2003.

XX 24-APR-2002; 2002US-00132067.

XX 24-APR-2002; 2002US-00132067.

XX (LALA-) LOS ALAMOS NAT LAB.

XX (REGC) UNIV CALIFORNIA.

XX Bradbury AM, Zeytun A, Waldo GS;

XX WPI; 2004-154325/15.

XX N-PSDB; ADI36420.

XX Novel binding ligand with intrinsic fluorescence and comprising
 PT fluorescent protein having heterologous binding sites, useful for
 PT detecting target molecule.

XX Claim 2; SEQ ID NO 4; 23pp; English.

XX The invention relates to binding ligands (fluorobodies) with intrinsic
 CC fluorescence, which comprises green fluorescent protein (GFP) having
 CC heterologous binding sites. The binding ligand is useful for detecting
 CC the target molecule and is efficiently detects the target molecule. The
 CC present sequence is Discosoma sp. red fluorescent protein (RED) used in
 CC the exemplification of the invention.

XX SQ Sequence 225 AA;

Query Match 100.0%; Score 1214; DB 8; Length 225;

Best Local Similarity 100.0%; Pred. No. 8.9e-128;

Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMECTVNGHEFEIEGEGRPEYGHNTVKLVKTKGGPLPFAWDI 60

DB 1 MRSSKNVKEFMRFKVRMECTVNGHEFEIEGEGRPEYGHNTVKLVKTKGGPLPFAWDI 60

QY 61 LSPQFOYGVSKVYVKKHPADIPDYKKLSPEGFKWERNVNFEDGGVVTQTSSLDQGCIFY 120

DB 61 LSPQFOYGVSKVYVKKHPADIPDYKKLSPEGFKWERNVNFEDGGVVTQTSSLDQGCIFY 120

QY 121 KVKFIGVNFPSDGPVNMQKTMGWEASTERLYPRDGLVKGEIHKALKDKGGHYLVPEFKSI 180

DB 121 KVKFIGVNFPSDGPVNMQKTMGWEASTERLYPRDGLVKGEIHKALKDKGGHYLVPEFKSI 180

QY 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTYVEQYTERGHRHLFL 225

DB 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTYVEQYTERGHRHLFL 225

Search completed: June 30, 2005, 15:46:34
 Job time : 165 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:43:54 ; Search time 161 Seconds
(without alignments)
539.155 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSSKNVKEFMFKVMEG.....EDYIVEQYVTEGRHHLFL 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 1722976

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubaa/US10E_PUBCOMB.pep.*
- 18: /cgn2_6/ptodata/2/pubaa/US10F_PUBCOMB.pep.*
- 19: /cgn2_6/ptodata/2/pubaa/US11_PUBCOMB.pep.*
- 20: /cgn2_6/ptodata/2/pubaa/US11_NEW_PUB.pep.*
- 21: /cgn2_6/ptodata/2/pubaa/US60_NEW_PUB.pep.*
- 22: /cgn2_6/ptodata/2/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	100.0	225	9	US-09-999-745-67
2	1214	100.0	225	10	US-09-866-538-12
3	1214	100.0	225	10	US-09-794-308-12
4	1214	100.0	225	10	US-09-865-291-12
5	1214	100.0	225	13	US-10-006-922-12
6	1214	100.0	225	14	US-10-081-864-8
7	1214	100.0	225	14	US-10-121-258-1
8	1214	100.0	225	14	US-10-315-920-2
9	1214	100.0	225	15	US-10-132-067-4
10	1214	100.0	225	15	US-10-370-570-56
11	1214	100.0	225	15	US-10-406-618-32

12	1214	100.0	225	16	US-10-433-640-13	Sequence 13, Appl
13	1214	100.0	225	16	US-10-311-030-7	Sequence 7, Appl
14	1214	100.0	225	16	US-10-845-484-3	Sequence 3, Appl
15	1214	100.0	225	16	US-10-885-988-12	Sequence 12, Appl
16	1214	100.0	225	17	US-10-656-029-22	Sequence 22, Appl
17	1214	100.0	225	17	US-10-857-623-12	Sequence 12, Appl
18	1214	100.0	225	17	US-10-505-486-26	Sequence 26, Appl
19	1214	100.0	487	15	US-10-343-977-1	Sequence 1, Appl
20	1214	100.0	487	17	US-10-343-977-1	Sequence 1, Appl
21	1214	100.0	506	15	US-10-343-977-2	Sequence 2, Appl
22	1214	100.0	506	17	US-10-343-977-2	Sequence 3, Appl
23	1214	100.0	547	15	US-10-343-977-3	Sequence 3, Appl
24	1214	100.0	547	17	US-10-343-977-3	Sequence 3, Appl
25	1210	99.7	226	16	US-10-311-030-9	Sequence 9, Appl
26	1210	99.7	240	14	US-10-152-296-2	Sequence 2, Appl
27	1210	99.7	240	16	US-10-739-656-2	Sequence 2, Appl
28	1210	99.7	545	14	US-10-214-932-52	Sequence 52, Appl
29	1210	99.7	548	14	US-10-214-932-76	Sequence 76, Appl
30	1207	99.4	225	14	US-10-121-258-20	Sequence 20, Appl
31	1207	99.4	225	14	US-10-315-920-4	Sequence 4, Appl
32	1196	98.5	236	16	US-10-314-936-2	Sequence 2, Appl
33	1196	98.5	236	16	US-10-314-936-4	Sequence 4, Appl
34	1196	98.5	236	19	US-11-021-014-2	Sequence 2, Appl
35	1196	98.5	236	19	US-11-021-014-4	Sequence 4, Appl
36	1195.5	98.5	226	16	US-10-724-178-12	Sequence 12, Appl
37	1191	98.1	225	14	US-10-315-920-6	Sequence 6, Appl
38	1186	97.7	225	15	US-10-442-148A-7	Sequence 7, Appl
39	1186	97.7	239	15	US-10-442-148A-8	Sequence 8, Appl
40	1160	95.6	225	14	US-10-121-258-4	Sequence 4, Appl
41	1153	95.0	225	14	US-10-121-258-24	Sequence 24, Appl
42	1132	93.2	225	15	US-10-423-688A-41	Sequence 41, Appl
43	1121	92.3	226	14	US-10-121-258-6	Sequence 6, Appl
44	1119	92.2	225	13	US-10-006-922-44	Sequence 44, Appl
45	1119	92.2	225	14	US-10-081-864-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-999-745-67

Query Match 100.0%; Score 1214; DB 9; Length 225;
Best Local Similarity 100.0%; Pred. No. 6, 1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MRSSKNVKEFMFKVMEGTVNGHFEIEGEGREYEGHNTVKLVTKGGPLPFAWDI	60
Db	1	MRSSKNVKEFMFKVMEGTVNGHFEIEGEGREYEGHNTVKLVTKGGPLPFAWDI	60
Qy	61	LSQFOYGVSKVYKHADIPDYKLSFPEGFKVNVNFDGGVTVTQSSLDGCFIY	120
Db	61	LSQFOYGVSKVYKHADIPDYKLSFPEGFKVNVNFDGGVTVTQSSLDGCFIY	120

Qy 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Db 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225

RESULT 2

US-09-866-538-12
; Sequence 12, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: TSJEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-866-538-12

Query Match 100.0%; Score 1214; DB 10; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVTKGGPLPPAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVTKGGPLPPAWDI 60
Qy 61 LSPQFQYGSKVYVYKHADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCIFY 120
Db 61 LSPQFQYGSKVYVYKHADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCIFY 120
Qy 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Db 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225

RESULT 3

US-09-794-308-12
; Sequence 12, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: TSJEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-794-308-12

Query Match 100.0%; Score 1214; DB 10; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;

Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVTKGGPLPPAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVTKGGPLPPAWDI 60
Qy 61 LSPQFQYGSKVYVYKHADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCIFY 120
Db 61 LSPQFQYGSKVYVYKHADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCIFY 120
Qy 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Db 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225

RESULT 4

US-09-865-291-12
; Sequence 12, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-865-291-12

Query Match 100.0%; Score 1214; DB 10; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVTKGGPLPPAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVTKGGPLPPAWDI 60
Qy 61 LSPQFQYGSKVYVYKHADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCIFY 120
Db 61 LSPQFQYGSKVYVYKHADIPDYKLSFPEGFKWERNVNFEDGGVTVTQDSSLQDGCIFY 120
Qy 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Db 121 KVKFIGVNFPSDGPVMQKKTGWGWEASTERLYPRDGLVKGELHKAHLKDGSHYLVFPKSI 180
Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225

RESULT 5

US-10-006-922-12
; Sequence 12, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A.
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; METHODS FOR USING THE SAME


```
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-12

Query Match
Best Local Similarity 100.0%; Score 1214; DB 13; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60

Qy 61 LSPQFYQYGSVYVVKHPADIPDYKKLSPEGFKWERNVNFEDGGVYVTVQSSLDQGCIFY 120
Db 61 LSPQFYQYGSVYVVKHPADIPDYKKLSPEGFKWERNVNFEDGGVYVTVQSSLDQGCIFY 120

Qy 121 KVKEFIGVNFPSDGPVMQKTMGEASTERLYPRDGLKGEIHKALKLDGGHYLVEFKSI 180
Db 121 KVKEFIGVNFPSDGPVMQKTMGEASTERLYPRDGLKGEIHKALKLDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 7
US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(225)
; OTHER INFORMATION: wild-type DeRed
US-10-121-258-1

Query Match
Best Local Similarity 100.0%; Score 1214; DB 14; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60

Qy 61 LSPQFYQYGSVYVVKHPADIPDYKKLSPEGFKWERNVNFEDGGVYVTVQSSLDQGCIFY 120
Db 61 LSPQFYQYGSVYVVKHPADIPDYKKLSPEGFKWERNVNFEDGGVYVTVQSSLDQGCIFY 120

Qy 121 KVKEFIGVNFPSDGPVMQKTMGEASTERLYPRDGLKGEIHKALKLDGGHYLVEFKSI 180
Db 121 KVKEFIGVNFPSDGPVMQKTMGEASTERLYPRDGLKGEIHKALKLDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 8
US-10-315-920-2
; Sequence 2, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; TITLE OF INVENTION: Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Query Match
Best Local Similarity 100.0%; Score 1214; DB 14; Length 225;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHFEIEGEGRPEYEGHNTVKLVKTKGGPLPFAWDI 60

Qy 61 LSPQFYQYGSVYVVKHPADIPDYKKLSPEGFKWERNVNFEDGGVYVTVQSSLDQGCIFY 120
Db 61 LSPQFYQYGSVYVVKHPADIPDYKKLSPEGFKWERNVNFEDGGVYVTVQSSLDQGCIFY 120

Qy 121 KVKEFIGVNFPSDGPVMQKTMGEASTERLYPRDGLKGEIHKALKLDGGHYLVEFKSI 180
Db 121 KVKEFIGVNFPSDGPVMQKTMGEASTERLYPRDGLKGEIHKALKLDGGHYLVEFKSI 180

Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
```

```
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; TITLE OF INVENTION: FOR THEIR USE
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; PRIOR FILING DATE: 2002-12-09
; PRIOR APPLICATION NUMBER: 60/211,607
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: PCT/US01/19097
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; US-10-315-920-2

Query Match      100.0%; Score 1214; DB 14; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
DB 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
QY 61 LSPQFOYGSKVYVKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTTQDSSLQDGCIFY 120
DB 61 LSPQFOYGSKVYVKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTTQDSSLQDGCIFY 120
QY 121 KVXFIGVNFPSDGPVMQKKTGMWEASTERLYPRDGLVKGKGIHAKLKDGGHYLVFVKSI 180
DB 121 KVXFIGVNFPSDGPVMQKKTGMWEASTERLYPRDGLVKGKGIHAKLKDGGHYLVFVKSI 180
QY 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225
DB 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225

RESULT 9
US-10-132-067-4
; Sequence 4, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
; TITLE OF INVENTION: Fluorescence
; FILE REFERENCE: 021362-000600US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; OTHER INFORMATION: red fluorescent protein (dsRED)
; US-10-132-067-4

Query Match      100.0%; Score 1214; DB 15; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
DB 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGEGRPYEGHNTVKLVKTKGGPLPFAWDI 60
QY 61 LSPQFOYGSKVYVKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTTQDSSLQDGCIFY 120
DB 61 LSPQFOYGSKVYVKHPADIPDYKLSFPEGFKWERNVNFEDGGVTVTTQDSSLQDGCIFY 120
QY 121 KVXFIGVNFPSDGPVMQKKTGMWEASTERLYPRDGLVKGKGIHAKLKDGGHYLVFVKSI 180
DB 121 KVXFIGVNFPSDGPVMQKKTGMWEASTERLYPRDGLVKGKGIHAKLKDGGHYLVFVKSI 180
QY 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225
DB 181 YMAKKPVQLPGYIYVDSKLDITSHNEDYTIQEYERTEGRHHLFL 225

RESULT 11
US-10-406-618-32
; Sequence 32, Application US/10406618
; Publication No. US20030219814A1
; GENERAL INFORMATION:
; APPLICANT: Wan, David Chi-Cheong
; APPLICANT: Ip, Denise Tsz-Ming
; APPLICANT: The Chinese University of Hong Kong
; TITLE OF INVENTION: No. US20030219814A1e1 Fluorescent Proteins
; FILE REFERENCE: 016285-34-1
; CURRENT APPLICATION NUMBER: US/10/406,618
; CURRENT FILING DATE: 2003-04-02
; PRIOR APPLICATION NUMBER: US 60/387,968
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: US 60/370,598
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
;
;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Spontaneously
; OTHER INFORMATION: Fluorescent protein Dered.
US-10-406-618-32

Query Match      100.0%; Score 1214; DB 15; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRYPYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRYPYEGHNTVKLVKTKGGPLPFAWDI 60

Qy 61 LSPQFQYGSVYVYKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
Db 61 LSPQFQYGSVYVYKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120

Qy 121 KVAFIGNFSDGPMVKMTGWEASTERLYPRDGLVKGHEIKALKDKGGHYLVPEFKSI 180
Db 121 KVAFIGNFSDGPMVKMTGWEASTERLYPRDGLVKGHEIKALKDKGGHYLVPEFKSI 180

Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 12
US-10-433-640-13
; Sequence 13, Application US/10433640
; Publication No. US20040115792A1
; GENERAL INFORMATION:
; APPLICANT: Lichtenberg-Frate, Hella
; TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
; TITLE OF INVENTION: COMPLEX ENVIRONMENTAL CONTAMINATIONS
; FILE REFERENCE: 1487/3
; CURRENT APPLICATION NUMBER: US/10/433,640
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/14610
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: DE 10061872.3
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-433-640-13

Query Match      100.0%; Score 1214; DB 16; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRYPYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRYPYEGHNTVKLVKTKGGPLPFAWDI 60

Qy 61 LSPQFQYGSVYVYKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
Db 61 LSPQFQYGSVYVYKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120

Qy 121 KVAFIGNFSDGPMVKMTGWEASTERLYPRDGLVKGHEIKALKDKGGHYLVPEFKSI 180
Db 121 KVAFIGNFSDGPMVKMTGWEASTERLYPRDGLVKGHEIKALKDKGGHYLVPEFKSI 180

Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 13
US-10-311-030-7
; Sequence 7, Application US/10311030
```

```
;
;
; Publication No. US20040171107A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamiaara, Elize
; APPLICANT: Teien, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/184,732
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-311-030-7

Query Match      100.0%; Score 1214; DB 16; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRYPYEGHNTVKLVKTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRYPYEGHNTVKLVKTKGGPLPFAWDI 60

Qy 61 LSPQFQYGSVYVYKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120
Db 61 LSPQFQYGSVYVYKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDQGCIFY 120

Qy 121 KVAFIGNFSDGPMVKMTGWEASTERLYPRDGLVKGHEIKALKDKGGHYLVPEFKSI 180
Db 121 KVAFIGNFSDGPMVKMTGWEASTERLYPRDGLVKGHEIKALKDKGGHYLVPEFKSI 180

Qy 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPVQLPGYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225

RESULT 14
US-10-845-484-3
; Sequence 3, Application US/10845484
; Publication No. US20040248180A1
; GENERAL INFORMATION:
; APPLICANT: Bulina, Maria
; APPLICANT: Chudakov, Dmitry
; APPLICANT: Lukanov, Konstantin
; TITLE OF INVENTION: Mutant Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Making and Using the Same
; FILE REFERENCE: CLON 092
; CURRENT APPLICATION NUMBER: US/10/845,484
; CURRENT FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: 60/343128
; PRIOR FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: PCT/US02/41418
; PRIOR FILING DATE: 2002-12-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-845-484-3

Query Match      100.0%; Score 1214; DB 16; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRPRYEGHNTVKLVTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRPRYEGHNTVKLVTKGGPLPFAWDI 60
Qy 61 LSPQFYGSKVYVKHPADI PDYKLSPEFGKWERVNFEDGGVTVTQDSSLQDGCIFY 120
Db 61 LSPQFYGSKVYVKHPADI PDYKLSPEFGKWERVNFEDGGVTVTQDSSLQDGCIFY 120
Qy 121 KVXFIGVNFPSDGPVMQKTMGEASTERLYPRDGLVKGHIKALKLDGGHYLVEFKSI 180
Db 121 KVXFIGVNFPSDGPVMQKTMGEASTERLYPRDGLVKGHIKALKLDGGHYLVEFKSI 180
Qy 181 YMAKKPQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
```

RESULT 15

```
US-10-885-988-12
; Sequence 12, Application US/10885988
; Publication No. US20040259165A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TS'EN, Roger
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/10/885,988
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-885-988-12
```

```
Query Match 100.0%; Score 1214; DB 16; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e-114;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRPRYEGHNTVKLVTKGGPLPFAWDI 60
Db 1 MRSSKNVKEFMRFKVRMEGTVNGHEFEIEGEGRPRYEGHNTVKLVTKGGPLPFAWDI 60
Qy 61 LSPQFYGSKVYVKHPADI PDYKLSPEFGKWERVNFEDGGVTVTQDSSLQDGCIFY 120
Db 61 LSPQFYGSKVYVKHPADI PDYKLSPEFGKWERVNFEDGGVTVTQDSSLQDGCIFY 120
Qy 121 KVXFIGVNFPSDGPVMQKTMGEASTERLYPRDGLVKGHIKALKLDGGHYLVEFKSI 180
Db 121 KVXFIGVNFPSDGPVMQKTMGEASTERLYPRDGLVKGHIKALKLDGGHYLVEFKSI 180
Qy 181 YMAKKPQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
Db 181 YMAKKPQLPGYYVYVDSKLDITSHNEDYTIIVEQYERTEGRHHLFL 225
```

Search completed: June 30, 2005, 15:53:58
Job time : 162 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2005, 15:41:04 ; Search time 43 Seconds
(without alignments)
390.606 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSSKNVKEFMRPKVMEG.....EDYIVQEYRTEGRHFL 225

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1210	99.7	240	4	US-10-152-296-2
2	485.5	40.0	238	3	US-09-277-716-16
3	485.5	40.0	238	4	US-09-609-161B-16
4	485.5	40.0	238	4	US-09-626-581D-65
5	485.5	40.0	238	4	US-09-415-765B-65
6	485.5	40.0	238	4	US-09-626-580C-65
7	473.5	39.0	224	4	US-09-977-897-3
8	473.5	39.0	224	4	US-09-977-897-13
9	473.5	39.0	225	4	US-09-977-897-14
10	473.5	39.0	226	4	US-09-977-897-15
11	473.5	39.0	227	4	US-09-977-897-16
12	473.5	39.0	228	4	US-09-977-897-17
13	473.5	39.0	229	4	US-09-977-897-18
14	473.5	39.0	230	4	US-09-977-897-19
15	473.5	39.0	231	4	US-09-977-897-20
16	473.5	39.0	232	4	US-09-977-897-21
17	473.5	39.0	233	4	US-09-977-897-22
18	473.5	39.0	234	4	US-09-977-897-9
19	473.5	39.0	234	4	US-09-977-897-23
20	473.5	39.0	235	4	US-09-977-897-8
21	473.5	39.0	235	4	US-09-977-897-24
22	473.5	39.0	236	4	US-09-977-897-7
23	473.5	39.0	236	4	US-09-977-897-25
24	473.5	39.0	237	4	US-09-977-897-6
25	473.5	39.0	237	4	US-09-977-897-26
26	473.5	39.0	238	3	US-09-277-716-32
27	473.5	39.0	238	4	US-09-609-161B-32

28	473.5	39.0	238	4	US-09-839-650-3	Sequence 3, Appli
29	473.5	39.0	238	4	US-09-977-897-5	Sequence 5, Appli
30	473.5	39.0	238	4	US-09-977-897-27	Sequence 27, Appli
31	473.5	39.0	239	4	US-09-977-897-2	Sequence 2, Appli
32	473	39.0	219	4	US-09-977-897-4	Sequence 4, Appli
33	473	39.0	231	4	US-09-977-897-12	Sequence 12, Appli
34	473	39.0	231	4	US-09-977-897-11	Sequence 11, Appli
35	473	39.0	233	4	US-09-977-897-10	Sequence 10, Appli
36	215.5	17.8	238	4	US-09-023-946B-23	Sequence 23, Appli
37	214.5	17.7	238	4	US-09-023-946B-28	Sequence 28, Appli
38	213.5	17.6	240	4	US-09-129-192C-49	Sequence 49, Appli
39	212.5	17.5	238	2	US-08-818-604-32	Sequence 32, Appli
40	212.5	17.5	238	3	US-08-819-612-22	Sequence 22, Appli
41	212.5	17.5	238	4	US-09-316-919-3	Sequence 3, Appli
42	212.5	17.5	238	4	US-09-346-946-32	Sequence 32, Appli
43	212.5	17.5	238	4	US-09-023-946B-22	Sequence 22, Appli
44	212.5	17.5	238	4	US-09-316-920A-3	Sequence 3, Appli
45	212.5	17.5	238	4	US-09-872-364-22	Sequence 22, Appli

ALIGNMENTS

RESULT 1
US-10-152-296-2

; Sequence 2, Application US/10152296
; Patent No. 6723537
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-000110US
; CURRENT APPLICATION NUMBER: US/10/152,296
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 2
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian
; OTHER INFORMATION: codon-optimized variant (dsRED) of Discosoma sp.
; OTHER INFORMATION: "red" red fluorescent protein (RFP)

US-10-152-296-2

Query Match	99.7%	Score	1210;	DB	4;	Length	240;
Best Local Similarity	99.6%	Pred. No.	5.7e-130;	Mismatches	0;	Indels	0;
Matches	224;	Conservative	1;				
Qy	1	MRSSKNVKEFMRPKVMEGTVNGHFEIEGEGRPVEGHNTVKLVTKGGPLPFAWDI	60				
Db	2	VRSSKNVKEFMRPKVMEGTVNGHFEIEGEGRPVEGHNTVKLVTKGGPLPFAWDI	61				
Qy	61	LSPOFYQSGYVVKHPADIIDYKLSFPEGPKWERNVNFEDGGVYVTTQDSSLQDGCIFY	120				
Db	62	LSPOFYQSGYVVKHPADIIDYKLSFPEGPKWERNVNFEDGGVYVTTQDSSLQDGCIFY	121				
Qy	121	KVKGFGVNFSDGVPVQKMTGWEASTERLYPRDGLVKGHIKALKDGGHYLVEPKSI	180				
Db	122	KVKGFGVNFSDGVPVQKMTGWEASTERLYPRDGLVKGHIKALKDGGHYLVEPKSI	181				
Qy	181	YMAKKPVQLGYYVDSKLDITSHNEDYTTVEQYRTEGRHHLFL	225				
Db	182	YMAKKPVQLGYYVDSKLDITSHNEDYTTVEQYRTEGRHHLFL	226				

RESULT 2
US-09-277-716-16
; Sequence 16, Application US/09277716A
; Patent No. 6232107

Qy 119 IYKVFQVGNFSPDGPVMQKTMGWEASTERLYPRDVLKGEIHKALKLDGGHYLVFEK 178
Db 122 VYRVEYKSGNSPDDGPVMQKTLGIEPSFEAMVMNGVLVGEVLVYKLSNGKIYYSCHMK 181
Qy 179 SIYMAKRPV-QLPGYVYVDSKLDITSHNEDYTIQEY 215
Db 182 TLMKSGVVKFPPSYHFIOHRLKLT-YVEDGGFVEQHE 218

RESULT 7
US-09-977-897-3
; Sequence 3, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-3

Query Match 39.0%; Score 473.5; DB 4; Length 224;
Best Local Similarity 45.5%; Pred. No. 8.3e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRKVRMEGTVNGHEPIEGEGEPYEGHNTVKLVKTKGGPLPFAWDILSP 63
Db 8 KNTGLKELMSKASVEGIVNNHVFSGEGFGKGNVLFGNOLMQIRVTYKGGPLPFAFDIVSI 67
Qy 64 QFYQSGKVVYVHPADIPDYKKLSFPPEGPKWERNMNFEDGGVVTVTODSSLDGCGFYIKVK 123
Db 68 AFQVGNRTFTKYPDADIYFVQSPFAGFFYERNLRFEDGAIVDIRSDISLEDDKPHYKVE 127
Qy 124 FIGVNFSPDGPVMQKTMGWEASTERLYPRDVLKGEIHKALKLDGGHYLVFEKSIYMA 183
Db 128 YRNGGFPNSGPMQKAILGMPSEFVWYVNMNSGLVGEVDLVYKLSNGKIYYSCHMKTFYRS 187
Qy 184 KKPV-QLPGYVYVDSKLDITSHNEDYTIQEY 215
Db 188 KGVKFPPEYHFIIHRLKLT-YVEGSGFVEQHE 219

RESULT 8
US-09-977-897-13
; Sequence 13, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-13

Query Match 39.0%; Score 473.5; DB 4; Length 224;
Best Local Similarity 45.5%; Pred. No. 8.3e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;


```

; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-17

Query Match          39.0%; Score 473.5; DB 4; Length 228;
Best Local Similarity 45.5%; Pred. No. 8.5e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRFKVMETGVNGHEFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSP 63
Db 8 KNTGLKEIMSAKASVEGIVNNHVFSEMGFGKGNVLFQNLQMQRVTKGGPLPFAFDIVSI 67

Qy 64 QFOYGSKVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDGCGFIYKVK 123
Db 68 AFQYGNRTFTKYPDDIADYFVQSPAGFFYERNLRFEDGAIVDIRSDISLEDDKFKHYKVE 127

Qy 124 FIGVNFPSDGPVMQKTMGWGEASERLYPRDGVLGKEIHKALKDKGCHYLVEFKSIYMA 183
Db 128 YRGNFGFNSGPMQKAILGMEPSFVYVNMNSGLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy 184 KKPV-QLPGYVYVDSKLDITSHNEDYTIIVEQYE 215
Db 188 KGGVKPEPEYHFHHRLEKT-YVEEGSFVEQHE 219

RESULT 13
US-09-977-897-18
; Sequence 18, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-18

Query Match          39.0%; Score 473.5; DB 4; Length 229;
Best Local Similarity 45.5%; Pred. No. 8.6e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRFKVMETGVNGHEFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSP 63
Db 8 KNTGLKEIMSAKASVEGIVNNHVFSEMGFGKGNVLFQNLQMQRVTKGGPLPFAFDIVSI 67

Qy 64 QFOYGSKVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDGCGFIYKVK 123
Db 68 AFQYGNRTFTKYPDDIADYFVQSPAGFFYERNLRFEDGAIVDIRSDISLEDDKFKHYKVE 127

Qy 124 FIGVNFPSDGPVMQKTMGWGEASERLYPRDGVLGKEIHKALKDKGCHYLVEFKSIYMA 183
Db 128 YRGNFGFNSGPMQKAILGMEPSFVYVNMNSGLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy 184 KKPV-QLPGYVYVDSKLDITSHNEDYTIIVEQYE 215
Db 188 KGGVKPEPEYHFHHRLEKT-YVEEGSFVEQHE 219

US-09-977-897-19
; Sequence 19, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-19

Query Match          39.0%; Score 473.5; DB 4; Length 230;
Best Local Similarity 45.5%; Pred. No. 8.7e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRFKVMETGVNGHEFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSP 63
Db 8 KNTGLKEIMSAKASVEGIVNNHVFSEMGFGKGNVLFQNLQMQRVTKGGPLPFAFDIVSI 67

Qy 64 QFOYGSKVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDGCGFIYKVK 123
Db 68 AFQYGNRTFTKYPDDIADYFVQSPAGFFYERNLRFEDGAIVDIRSDISLEDDKFKHYKVE 127

Qy 124 FIGVNFPSDGPVMQKTMGWGEASERLYPRDGVLGKEIHKALKDKGCHYLVEFKSIYMA 183
Db 128 YRGNFGFNSGPMQKAILGMEPSFVYVNMNSGLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy 184 KKPV-QLPGYVYVDSKLDITSHNEDYTIIVEQYE 215
Db 188 KGGVKPEPEYHFHHRLEKT-YVEEGSFVEQHE 219

RESULT 15
US-09-977-897-20
; Sequence 20, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-20

Query Match          39.0%; Score 473.5; DB 4; Length 231;
Best Local Similarity 45.5%; Pred. No. 8.7e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRFKVMETGVNGHEFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSP 63
Db 8 KNTGLKEIMSAKASVEGIVNNHVFSEMGFGKGNVLFQNLQMQRVTKGGPLPFAFDIVSI 67
```

```

Db 188 KGGVKPEPEYHFHHRLEKT-YVEEGSFVEQHE 219

RESULT 14
US-09-977-897-19
; Sequence 19, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-19

Query Match          39.0%; Score 473.5; DB 4; Length 230;
Best Local Similarity 45.5%; Pred. No. 8.7e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRFKVMETGVNGHEFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSP 63
Db 8 KNTGLKEIMSAKASVEGIVNNHVFSEMGFGKGNVLFQNLQMQRVTKGGPLPFAFDIVSI 67

Qy 64 QFOYGSKVYVVKHPADIPDYKLSPEGFKWERNVNFEDGGVVTVTQSSLDGCGFIYKVK 123
Db 68 AFQYGNRTFTKYPDDIADYFVQSPAGFFYERNLRFEDGAIVDIRSDISLEDDKFKHYKVE 127

Qy 124 FIGVNFPSDGPVMQKTMGWGEASERLYPRDGVLGKEIHKALKDKGCHYLVEFKSIYMA 183
Db 128 YRGNFGFNSGPMQKAILGMEPSFVYVNMNSGLVGEVDLVYKLESGNYYSCHMKTFYRS 187

Qy 184 KKPV-QLPGYVYVDSKLDITSHNEDYTIIVEQYE 215
Db 188 KGGVKPEPEYHFHHRLEKT-YVEEGSFVEQHE 219

RESULT 15
US-09-977-897-20
; Sequence 20, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-20

Query Match          39.0%; Score 473.5; DB 4; Length 231;
Best Local Similarity 45.5%; Pred. No. 8.7e-46;
Matches 97; Conservative 40; Mismatches 73; Indels 3; Gaps 3;

Qy 5 KNV-IKEFMRFKVMETGVNGHEFEIEGEGRYPYEGHNTVKLVTKGGPLPFAWDILSP 63
Db 8 KNTGLKEIMSAKASVEGIVNNHVFSEMGFGKGNVLFQNLQMQRVTKGGPLPFAFDIVSI 67
```

```

Qy 64 QFQYGSKYVVKHPADI PDYKKLSPEGFKWERVMNFEDGGVVTVTQDSSLQDGCFIYKVK 123
Db 68 AFQYGNFTTKYPDDIADYFVQSPAGFFYERNLRFEDGALVDIRSDISLEDDKFKHYKVE 127
Qy 124 FIGVNFPSDGPVMOKTWGWEASTERLYPRDGVLKGBIHKALKJDKGGHYLVBFKSIYMA 183
Db 128 YRGNGFPSPNGPVMQKAILGMEPSFEVVMNSGVLVGEVDLVYKLESGNYISCHMKTFYRS 187
Qy 184 KKPV-QLPGYVYVDSKLDITSHNEDYTIQEY 215
Db 188 KGVKEPPEYHFHHRLEKT-YVEEGSFVEQHE 219

```

Search completed: June 30, 2005, 15:51:10
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 1, 2005, 18:20:13 ; Search time 3340 Seconds
(without alignments)
2564.210 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSKKNVKEPMFKVRMEG.....EDYIVEQYRTSGRHHFL 225

Scoring table:

	BLOSUM62	Xgapop 10.0	Xgapext 0.5
Ygapop 10.0			0.5
Ygapop 6.0			7.0
Deiop 6.0			Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2/1/USPTO/spool/US10006922/runat_30062005_110616_19095/app_query.fasta_1.391
-DB-EST -QFMT-fastcap -SUFFIX=rst -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US10006922 @CGN 1 1 3437 @runat_30062005_110616_19095 -ICPU=3
-NO MMAP -LARGOQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*	1: gb_est1.*
2: gb_est2.*	
3: gb_hcc.*	
4: gb_est3.*	
5: gb_est4.*	
6: gb_est5.*	
7: gb_est6.*	
8: gb_gsl.*	
9: gb_gsl2.*	

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225	18.5	661	4	BI385436
2	216.5	17.8	637	4	BI379468
3	216.5	17.8	687	4	BI387917
4	212.5	17.5	643	4	BI386699
5	212.5	17.5	675	4	BI382638
6	209.5	17.3	790	1	AL044652
7	198.5	16.4	613	4	BI386976
8	192.5	15.9	629	4	BI388023
9	190	15.7	681	4	BI386888

10	188	15.5	629	4	BI384961
11	187	15.4	707	4	BI382442
12	185.5	15.3	689	4	BI380055
13	184.5	15.2	571	4	BI387406
14	180.5	14.9	550	5	BQ551001
15	175	14.4	671	4	BI381780
16	173.5	14.3	561	4	BI379061
17	172.5	14.2	577	4	BI386955
18	168.5	13.9	591	4	BI382373
19	163	13.4	563	4	BI382999
20	161.5	13.3	459	4	BI386952
21	158.5	13.1	537	4	BI376909
22	131	10.8	437	4	BI388504
23	128	10.5	453	4	BI385638
24	127	10.5	629	4	BI380008
25	122	10.0	617	4	BI381719
26	118	9.7	541	4	BI380627
27	111	9.1	527	4	BI381765
28	110.5	9.1	511	4	AJ437840
29	110	9.1	397	4	BI384677
30	107.5	8.9	675	6	CA352443
31	104.5	8.6	457	4	BI378182
32	102.5	8.4	397	1	AJ437938
33	101.5	8.4	325	4	BI386240
34	101.5	8.4	531	4	BI376416
35	98.5	8.1	317	1	AJ437931
36	98	8.1	1302	8	CC300765
37	97.5	8.0	633	4	BI376444
38	95	7.8	697	5	BU626888
39	94.5	7.8	489	4	BI381509
40	94	7.7	553	5	EX908917
41	93	7.7	921	2	BF312529
42	92.5	7.6	1573	3	CR656136
43	92	7.6	910	7	CN320807
44	92	7.6	930	5	BQ939912
45	92	7.6	968	9	CG771777

ALIGNMENTS

RESULT 1

LOCUS

BI385436

DEFINITION

BFL26_000139 Amphioxus 26hr cdna library (Name convention: BFL26 or

MPMG531) Branchiostoma floridae cDNA clone MPMG531H0446 5', mRNA

sequence.

ACCESSION

BI385436

VERSION

BI385436.1

KEYWORDS

EST.

SOURCE

Branchiostoma floridae (Florida lancelet)

ORGANISM

Branchiostoma floridae

Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

Branchiostoma.

REFERENCE

1 (bases 1 to 661)

AUTHORS

Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,

Herwig, R., Vingron, M. and Lehrach, H.

TITLE

New evidence for genome-wide duplications at the origin of

vertebrates using an amphioxus gene set and completed animal

genomes

JOURNAL

Genome Res. 13 (6A), 1056-1066 (2003)

MEDLINE

22683279

PUBMED

12799346

COMMENT

Contact: Panopoulou G

Laboratory 145, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr. 63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoulou@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGCGCTTACACTTTATGCTCCGCGTG 3' (M13RSP)
BACKWARD: 5' GCTATTACCGCTAGCTGGCAAGGGGATGTG 3' (M13FSP)

Insert Length: 1 Std Error: 0.00

Seq primer: 5'-CCGCTCCGGAATCCCGGT-3' pSport3/86

High quality sequence stop: 661.

FEATURES
SOURCE

```

1. .661
source
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531H0446"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 26hr cdNA lib
BFL26 or MPMGp531"
/notes="Vector: pSport1 (Gibco BRL);
ECORI (5'); Site 2: NotI, BamHI, H
primed and directionally cloned in
NotI (5'- PGACTAGTTCTAGATCGGAGCGCC
Sall 5'- TCGACCCACGGCTCG-3' adapter
ORIGIN

Alignment Scores:
Pred. No.: 1,77e-16
Score: 225.00
Percent Similarity: 47.76%
Best Local Similarity: 30.85%
Query Match: 18.53%
DB: 4

US-10-006-922A-12 (1-225) x B1385436 (1-661)

```

US-10-006-922A-12 (1-225) x B1395436 (1-661)

QY	16	ValArgMetGluGlyThrValAsnGlyHisGluPheGluLeuGluGlyGluGlyGluGly	35
Db	65	ATCCACCTCCACGGCTCCATCAAGGGCCACGAGTTCCAGCTGCTGGCGCGGAGCAGCG	124
QY	36	ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro	55
Db	125	GACCCGAGCGCGCTCCCTGGTGACACGCGAAATCCACCAAG--GGTCCCTCTGAAG	181
QY	56	PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis	75
Db	182	TTCTCTCCCTACTTGATGATCCCCACCTCCGGGTACGGGTACTTACCAGTACCTCCCTAC	241
QY	76	ProAlaAspIleProAspTyrLysValLysSerPheProGlu-----GlyPheLysTrp	93
Db	242	CCG---GACGGACCCCTCGCCTTTCCAGGCCACCATGTTGGAAGGATCAGGGTATGCGATC	298
QY	94	GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu	113
Db	299	CATCGGTGTCGACTTTGAGACGGAGGCCAAGCTGTCATCGAGTTTAAATACTCTAC	358
QY	114	GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly	133
Db	359	GAGGTTTCCCATATCAAGCGCCGACATGAAGTTCAAGGGAAACCGGTTTCTCGAGACGGG	418
QY	134	ProValMetGlnLysLysThrMetClyTrpGluAlaSer---ThrGluArgLeuTyrPro	152
Db	419	CCGGTCTATGACCGCATGTTGCGACAGCAGCGGCTGCGTGCTCCAAAGAACCTTACCT	478
QY	153	ArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHis	172
Db	479	AACGACAAACCATCGTCGACAGCTTCGACTGGACTTTACAACCTGCAGATGGGAGCGC	538

Qy	173	TyrLeuValGluPheLysSerIleTyr- :::	-----Met 182
Db	539	TACAAGGCCCAAGTCAGACCACTACATCTTCGGCAAGCCCTTTCAGCGCATCTCATG	598
Qy	183	AlaLysLysProValGlnLeuProGlyTyrTyrTyrValAsp---SerLysLeuAspIle	201
Db	599	AGAAGACAGCCGGTCTTCGTGTATCCGCAAGTGTCTACGTGAACGCTTCCAAAGACCGAATC	658
Qy	202	Thr 202 	
Db	659	ACC 661 	
RESULT 2			
LOCUS	BI379468		
DEFINITION	BFLG1_001099 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGp498) Branchiostoma floridae cDNA clone MPMGp49801918 5', mRNA sequence.	637 bp mRNA linear EST 26-AUG-2003	
ACCESSION	BI379468		
VERSION	BI379468.1	GI:30914617	
KEYWORDS	EST.		
SOURCE	Branchiostoma floridae (Florida lancelet)		
ORGANISM	Branchiostoma floridae		
REFERENCE	1 (bases 1 to 637)		
AUTHORS	Panopoulou,G., Hennig,S., Groth,D., Krause,A., Foustka,A.J., Herwig,R., Vingron,M. and Lehrach,H.		
TITLE	New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes		
JOURNAL	Genome Res. 13 (6A), 1056-1066 (2003)		
MEDLINE	22683279		
PUBMED	12799346		
COMMENT	Contact: Panopoulou G Laboratory 145, dept Lehrach Max-Planck-Institut fuer Molekulare Genetik Innestr.63-73, D-14195 Berlin, Germany Tel: +49 30 8413 1235 Fax: +49 30 8413 1128 Email: panopoul@molgen.mpg.de The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: http://www.molgen.mpg.de/amphioxus/ Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (http://www.rzpd.de) PCR Primers FORWARD: 5' CCCAGCGTTTACATTATGCTTCGGCTCG 3' (M13RSP) BACKWARD: 5' GCTATTAGCCAGCTGGCGAAAGGGGATGTG 3' (M13FSP) Insert Length: 1200 Std Error: 0.00 Seq primer: 5'-CCGTCGCGAATTCCTCCGGGT-3' pSport3/86 High quality sequence stop: 637. Location/Qualifiers 1. .637 /organism="Branchiostoma floridae" /mol_type="mRNA" /db_xref="taxon:7739" /clone="MPMGp49801918" /tissue_type="whole embryo" /dev_stage="5-6 hrs (gastrula stage)" /lab_host="E.coli, XL1 blue" /clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGp498)" /notes="Vector: pSport1; Site 1: Sali, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and		
FEATURES			
source			


```

Qy 114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly 133
Db 393 GAGGGTTCCCATATCAAGGCCGACATGAGTTACGGGAAACCGGTTTCCCTGAGGACGG 452
Qy 134 ProValMetGlnLysLysThrMetGlyTyrTrpGluAlaSer---ThrGluArgLeuTyrPro 152
Db 453 CCGGTTCATGACGACGACGAGATTGTCGACGAGCGGCTGCGTGTCCAGAAACACCTACCTT 512
Qy 153 ArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHis 172
Db 513 AACGACAAACACCATCTGTTGGACAGCTTCGACGTGACCTGACAACTGCAAGATGGGAAGCG 572
Qy 173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
Db 573 TACAAGCCCAAGTGAAGGACGACCTACATCTTCGCGCAAGCCC 614

RESULT 4
BI386699
LOCUS
DEFINITION
BFL26 001402 Amphioxus 26hr cDNA library (Name convention: BFL26 or
MPMGp531) Branchiostoma floridae cDNA clone MPMGp531L1370 5', mRNA
sequence.
ACCESSION
BI386699.1 GI:30921604
VERSION
BI386699.1
KEYWORDS
EST.
SOURCE
Branchiostoma floridae (Florida lancelet)
ORGANISM
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
1 (bases 1 to 643)
Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,
Herwig,R., Vingron,M. and Lehrach,H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
laboratory 145, dept Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (http://www.rzpd.de)
PCR Primers
FORWARD: 5' CCCGAGCTTACATTTATGCTTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCGCAGCTGGGAAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATTCCTCCGGGT-3' pSport3/86
High quality sequence stop: 643.
FEATURES
source
Location/Qualifiers
1..643
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531L1370"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neural stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention:

```

```

BFL26 or MPMGp531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligo dT
primed and directionally cloned in pSport1 vector using a
NotI (5'-TGACTAGTTCTAGATCGAGCGGCCGCC (T)15-3' and a
SalI 5'-TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN
Alignment Scores: 5.58e-15 Length: 643
Pred. No.: 212.50 Matches: 56
Score: 47.87% Conservative: 34
Percent Similarity: 29.79% Mismatches: 87
Best Local Similarity: 17.50% Indels: 11
Query Match: 4 Gaps: 5
DB: 4

US-10-006-922A-12 (1-225) x BI386699 (1-643)
Qy 16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGly 35
Db 81 ATCCACCTTCACGCTCCATCAACGCCACGAGTTCGACATGTTGGGGGAGGAAAGGC 140
Qy 36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
Db 141 GACCCGAAACGCGGCTCGCTGGTGACACAGCAAAATCCACCAAG---GGTCCCTGAAG 197
Qy 56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
Db 198 TTCTCTCCCTTACTGATGATCCCCACCTCGGGTACGGGTACTACCACTACCTCCCTAC 257
Qy 76 ProAlaAspIleProAspTyrLysLysLeuSerPheProGlu-----GlyPheLysTrp 93
Db 258 CCG---GAGGACCCCTCGCTTTCAGGCTCCATGTTGGAAGGATCGGGGTATGCAGTC 314
Qy 94 GluArgValMetAsnPheGluLysGlyGlyValValThrValThrGlnAspSerSerLeu 113
Db 315 TACCGGTGTTCGATTTGAAGACGAGGCAAGCTGACTACCGAGTTTAAGTACTCTCTAC 374
Qy 114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly 133
Db 375 GAGGGTTCCTATATCAAGCCGACATGAAGCTGATGGGAAGCGGTTTCCCTGACGAGCGC 434
Qy 134 ProValMetGlnLysLysThrMetGlyTyrTrpGluAlaSer---ThrGluArgLeuTyrPro 152
Db 435 CCAGTCATGACACGACGAGATTGTCGAGGACGCGGTGCGTCCCAAGAAGACGATCTCT 494
Qy 153 ArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLysAspGlyGlyHis 172
Db 495 AACAAACACACCATCTCGTGGACAGCTTCGACTGGAGTTACAACTGAGATGGGAAGCGC 554
Qy 173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro----- 186
Db 555 TACAGGCGCGAGTGTCTNAGCCACTACATCTTCGACAGCCCTTTTCAGCCGATCTCATG 614
Qy 187 ValGlnLeuProGlyTyrTyrTyr 194
Db 615 ATAAAGCAGCGCGTCTTCGTGTAC 638

RESULT 5
BI382638
LOCUS
DEFINITION
BFLG2 000815 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMGp498) Branchiostoma floridae cDNA clone MPMGp498F0258 5',
mRNA sequence.
ACCESSION
BI382638
VERSION
BI382638.1 GI:30917704
KEYWORDS
EST.
SOURCE
Branchiostoma floridae (Florida lancelet)
ORGANISM
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
1 (bases 1 to 675)
Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J.,

```

Herwig, R., Vingron, M. and Lehrach, H.
 New evidence for genome-wide duplications at the origin of
 vertebrates using an amphioxus gene set and completed animal
 genomes
 Genome Res. 13 (6A), 1056-1066 (2003)
 22683279
 MEDLINE
 PUBMED
 COMMENT
 Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihnestr. 63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoulou@molgen.mpg.de
 The library was characterised by oligonucleotide fingerprinting
 (ONFP) to reduce sequencing redundancy. According to the ONFP
 procedure, clones giving the same hybridisation pattern with a
 battery of 200 8mer oligonucleotides are grouped into clusters. One
 clone per cluster is selected for sequencing. The size of each
 cluster is an indicator of the frequency of a transcript in the
 analysed library. The cluster size as well the coordinates of the
 rest of the clones assigned to the same fingerprint cluster as the
 clone from which the above EST is generated is available at the
 amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary
 Database of the German Genome Project (<http://www.rzpd.de>)
 PCR Primers
 FORWARD: 5' CCCAGGCTTTACATTTATGCTTCGGGCTCG 3' (M13RSP)
 BACKWARD: 5' GCTATTACGCGACGTCGCGAAGGGGGGATGTG 3' (M13FSP)
 Insert Length: 1 Std Error: 0.00
 Seq primer: 5'-CCGCTCGGAATTCCTCCGGT-3' pSport3/86
 High quality sequence stop: 675.

FEATURES

Location/Qualifiers
 1..675
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPGp498F0258"
 /tissue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XL1 blue"
 /clone_lib="Amphioxus 5-6 hrs cDNA library (Name
 convention: BFLG or MPGp498)"
 /notes="vector: pSport1; Site_1: SalI, KpnI, EcoRI (5');
 Site_2: NotI, BamHI, HindIII (3'); OligodT primed and
 directionally cloned in pSport1 vector using a NotI
 (5'-pGACTAGTTCTAGATCGGCGCGCC (T)15-3' and a SalI 5'-
 TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Alignment Scores:
 Pred. No.: 5.98e-15 Length: 675
 Score: 212.50 Matches: 54
 Percent Similarity: 49.43% Conservative: 32
 Best Local Similarity: 31.03% Mismatches: 83
 Query Match: 17.50% Indels: 5
 DB: 4 Gaps: 4
 US-10-006-922A-12 (1-225) x B1382638 (1-675)
 Qy 16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluLeuGluGlyGluGly 35
 Db 110 ATCCACCTTCACGGCTCCATACGCGCCACGAGTTCGACATGTTGGGGGAGAAAGGC 169
 Qy 36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
 Db 170 GACCCGAACGCGGCTCGTGGTGACACGCGAATCCACCAAG---GGTCCCTGAG 226
 Qy 56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
 Db 227 TTCTCCCTTACTTGATGATCCCTCCACCTCGGTACGGGTACTACAGTACCTCCCTAC 286
 Qy 76 ProAlaAspIleProAspTyrLysLysLeuSerPheProGlu-----GlyPheLysTyr 93

Db 287 CCG---GACGAGCCCTCGCCTTCCAGGTCTCCATGTTGAAGATCGGGGTATGACGTC 343
 Qy 94 GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
 Db 344 TACCGGTGTTTCGACTTTGAAGACGAGGCAAGCTGACTACCGAGTTTAAGTACTCTCTAC 403
 Qy 114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly 133
 Db 404 GAGGGTTCCATATCAAGCCGACATGAAGCTGATGGGAAGCGGTTTTCCCTGACGAGCGC 463
 Qy 134 ProValMetGlnLysLysThrMetGlyTyrGluAlaSer--ThrGluArgLeuTyrPro 152
 Db 464 CCAGTATGACCCAGCCAGATTGTCACGACGCGGTGCGTGTCCAGAGAAGACGTATCTT 523
 Qy 153 ArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLysAspGlyGlyHis 172
 Db 524 AACACAACACCATCTGGGACAGCTTCGAGTGGAGTTACAACTGCAGATGGAAGCGC 583
 Qy 173 TyrLeuValGluPheLysSerIleTyrMetAlaLysPro 186
 Db 584 TACAAGGCCCGAGTGTCTNAGCCACTACATCTTCGACAAAGCCC 625
 RESULT 6
 AL044652 790 bp mRNA linear EST 04-SEP-2003
 LOCUS DKFZP434P092.r1 434 (synonym: htes3) Homo sapiens cDNA clone
 DEFINITION DKFZP434P092 5', mRNA sequence.
 ACCESSION AL044652
 VERSION AL044652.1 GI:5432867
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 790)
 AUTHORS Ansoorge,W., Benes,V., Krieger,S., Mewes,H.W., Gassenhuber,J. and
 Wiemann,S.
 TITLE EST (Ansoorge, Benes, et al.)
 JOURNAL Unpublished (1999)
 COMMENT Contact: MIPS
 MIPS
 Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by EMBL (European Molecular Biology Laboratories,
 Heidelberg/Germany) within the cDNA sequencing consortium of the
 German Genome Project.
 No 5' sequence available.
 This clone (DKFZP434P092) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 FEATURES
 source
 1..790
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZP434P092"
 /tissue_type="testis"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="434 (synonym: htes3)"
 /notes="vector: pSport1; Site_1: NotI; Site_2: SalI"

ORIGIN

Alignment Scores:
 Pred. No.: 1.73e-14 Length: 790
 Score: 209.50 Matches: 55
 Percent Similarity: 47.25% Conservative: 48
 Best Local Similarity: 25.23% Mismatches: 102
 Query Match: 17.26% Indels: 13
 DB: 1 Gaps: 6


```

QY 134 ProValMetGlnLysLysThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro 152
Db 446 CCAGTCATGACCCAGCAGGATTCGACACGAGCGGCTCGGTGTCACCAAGAGCATGATCTT 505

QY 153 ArgAspGlyValLeuLysGlyGluHisLysLeuLysLysLeuLysAspGlyGlyHis 172
Db 506 AACACAAACACCATCGTGGACAGCTTCGACTGGAGTTACAACCTGCAGATGGGAGCGC 565

QY 173 TyrLeuValGluPheLysSerIleTyrMetAlaLysLysProValGln 188
Db 566 TACAGGGCCCGAGTGTCCGAGCAGCTACATCTTCACAGCGCCCTTCAA 613

RESULT 8
LOCUS BI388023 629 bp mRNA linear EST 26-AUG-2003
DEFINITION BFL26 002726 Amphioxus 26hr cdNA library (Name convention: BFL26 or
MPMGp531) Branchiostoma floridae cdNA clone MPMGP531J1737 5', mRNA
sequence.
ACCESSION BI388023.1 GI:30922844
VERSION BI388023
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
REFERENCE 1 (bases 1 to 629)
AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J.,
Herwig, R., Vingron, M. and Lehrach, H.
TITLE New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE 22683279
PUBMED 12799346
COMMENT Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (http://www.rzpd.de)
PCR Primers
FORWARD: 5' CCCAGCTTTACATTTATGCTTCGGGCTCG 3' (M13RSP)
BACKWARD: 5' GTATTACCGCAGCTCGGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86
High quality sequence stop: 629.
Location/Qualifiers
1. 629
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531J1737"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, Xli blue"
/clone_lib="Amphioxus 26hr cdNA library (Name convention:
BFL26 or MPMGP531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT
primed and directionally cloned in pSport1 vector using a

```

```

ORIGIN
Alignment Scores:
Pred. No.: 1,44e-12 Length: 629
Score: 192.50 Matches: 53
Percent Similarity: 47.19% Conservative: 31
Best Local Similarity: 29.78% Mismatches: 82
Query Match: 15.86% Indels: 12
DB: 4 Gaps: 4

US-10-006-922A-12 (1-225) x BI388023 (1-629)
QY 16 ValArgMetGluGlyThrValAenGlyHisGluPheGluLeuGluGlyGluGly 35
Db 83 ATCCACCTCCACGGCTCCATCAACGGGACGAGTTCGACCTGGTGGCGGAGGACAAGGC 142
QY 36 ArgProTyrGluGlyHisThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
Db 143 GACCCGAACCGCGGCTCGTGGTGACACAGAGAAATCCACCAAG---GGTCCCTGAAG 199
QY 56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
Db 200 TTCTCTCCCTACTGTATGATCCCGGCTGGTACGGGTACTACCACTACCTCCCTAC 259
QY 76 Pro-AlaAspIleProAspTyrLysLysLeuSerPhePro-----GluG 90
Db 260 CCGGACGGAC-----CCTCGCCTTTCCAGGCCACCATGTTGGGAGGATCAGGG 307
QY 90 YPheLysTrpGluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAs 110
Db 308 GTATGAGTCCATCGCGTGTTCGACTTTGAACGCGGAGGACGCTGTCCATCGAGTTAA 367
QY 110 pSerSerLeuGlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPhePr 130
Db 368 ATACTCTACGAGGGTTCCTATATCAAGCGGCAGATGAAGTTACGGGAACCGGTTTCCC 427
QY 130 oSerAspGlyProValMetGlnLysLysThrMetGlyTyrGluAlaSer---ThrGluAr 149
Db 428 TGATGACGGGCGGTCTATGACACGACGAGATTGTCTACCAACGCGGCTCGCTGTCCAAGAA 487
QY 149 gLeuTyrProArgAspGlyValLeuLysGlyGluHisLysAlaLeuLysLeuLysAs 169
Db 488 GACCTAGCTTTAACGACAAGACCATCTGGGACAGCTTCGACTGGCATTTACAACCTGCACAA 547
QY 169 pGlyGlyHisTyrLeuValGluPheLysSerIleTyrMetAlaLysLysPro 186
Db 548 TGGGAAGCGCTACAAGGGCCAGTACGAGCCACTACATCTTCGGGAAGGCC 599

RESULT 9
LOCUS BI386888 681 bp mRNA linear EST 26-AUG-2003
DEFINITION BFL26 001591 Amphioxus 26hr cdNA library (Name convention: BFL26 or
MPMGp531) Branchiostoma floridae cdNA clone MPMGP531119122 5', mRNA
sequence.
ACCESSION BI386888
VERSION BI386888.1 GI:30921785
KEYWORDS EST.
SOURCE Branchiostoma floridae (Florida lancelet)
ORGANISM Branchiostoma floridae
REFERENCE 1 (bases 1 to 681)
AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J.,
Herwig, R., Vingron, M. and Lehrach, H.
TITLE New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE 22683279
PUBMED 12799346
COMMENT Contact: Panopoulou G

```

laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>.
Database and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers

FORWARD: 5' CCCAGGCTTTACACTTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GTATTACCCAGCTGGCGAAAGGGGATGTG 3' (M13FSP)

Insert Length: 1 Std Error: 0.00

Seq primer: 5'-CCGTCGGGAATCCCGGT-3' pSport3/86

High quality sequence stop: 681.

Location/Qualifiers

1. .681

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMG531119122"

/tissue_type="whole embryo"

/dev_stage="26 hrs (neuroblast stage)"

/lab_host="E.coli, XL1 blue"

/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531)"

/notes="vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligo primed and directionally cloned in pSport1 vector using a NotI (5')-PGACTAGTTCTAGATCGGCGGCCGCC (T)15-3' and a SalI 5'-TCGACCCACGGCTCGG-3' adapters (Gibco BRL)."

ORIGIN

Alignment Scores:

Pred. No.:	3.24e-12	Length:	681
Score:	190.00	Matches:	53
Percent Similarity:	45.18%	Conservative:	36
Best Local Similarity:	26.90%	Mismatches:	98
Query Match:	15.65%	Indels:	10
DB:	4	Gaps:	4

US-10-006-922A-12 (1-225) x BI386888 (1-681)

QY	20	GlyThrValAsnGlyHisGluPheGluLeuGlyGlyGlyGlyGlyArgProTyrGlu	39
DB	98	GGCTCGGTCAATCGCGGAGTTCGACCTGGTGGAGCGGAAAGGCAACCGAAGCAT	157
QY	40	GlyHisAsnThrValLysLeuValThrLysGlyGlyProLeuPheAlaTyrAsp	59
DB	158	GGAACGCTCAGACCATGAGTGAATCCACCCGGGGGCC---CTGCCCTGCTCCCGCTG	214
QY	60	IleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIle	79
DB	215	CTGATCGGACCAACCTGGGGTACGGCTTACACGATACCTGCTCCCTGGCGGGCC	274
QY	80	ProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsn	99
DB	275	TCACCTCTCCAAACCGCCATCATCGGACGGAGGTACAGGTTCCACCGTGTTCAGATTT	334
QY	100	GluAspGlyGlyValValThrValThrGlnAspSerLeuGlnAspGlyCysPheIle	119
DB	335	GAGAGCGCGGAGTGTCTCAATTCGAACTCCGCTACACCTACAGGCGGCAAGATCAA	394
QY	120	TyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLys	139

Db	395	GGGAGTTCAGCTGATCGGTTCAGTTTCCTCGCGGGTCCCGTGATGTCGGCGGA	454
QY	140	ThrMetGlyTrpGluAlaSerThrGluArgLeuTyr---ProArgAspGlyValLeuLys	158
Db	455	CTGACCACCTCGGACAGGAGCGTGGCCAACTGCAGTGTCTCGACACCGCACCATCACC	514
QY	159	GlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLys	178
Db	515	GGCACTAACAGTGGAGCTTCGCACACCGATGGGAAGCGGTATCAGGGGACGTGCAG	574
QY	179	SerIleTyrMetAlaLysLysPro-----ValGlnLeuProGlyTyr	192
Db	575	ACGAACCTACACCTTCGCCAAGCGCTCCCGCGGTCTCATGTAAAGATCGCG-----	628
QY	193	TyrTyrValAspSerLysLeuAspIleThrSerHisAsnGluAspTyrThr	209
Db	629	ATCTTCTGGGCAACAGATCGAGGTCAAGGCGTCCACGACCGACATCACC	679

RESULT 10

BI384961

LOCUS

DEFINITION

BI384961.1 GI:30919957

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PubMed

COMMENT

BI384961 629 bp mRNA linear EST 26-AUG-2003
BFLG2_0031138 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498E05107 5', mRNA sequence.

BI384961.1 GI:30919957

EST.

Branchiostoma floridae (Florida lancelet)

Branchiostoma floridae

Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.

1 (bases 1 to 629)

Panopoulou,G., Hennig,S., Groth,D., Krause,A., Foustka,A.J.,

Herwig,R., Vingron,M. and Lehrach,H.

New evidence for genome-wide duplications at the origin of

vertebrates using an amphioxus gene set and completed animal

genomes

Genome Res. 13 (6A), 1056-1066 (2003)

22683279

12799346

Contact: Panopoulou G

Laboratory 145, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr. 63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each

cluster is an indicator of the frequency of a transcript in the

analysed library. The cluster size as well the coordinates of the

rest of the clones assigned to the same fingerprint cluster as the

clone from which the above EST is generated is available at the

amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>.

Database and filters are distributed via the Resource Center/Primary

Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGGCTTTACACTTTATGCTTCGGCTCG 3' (M13RSP)

BACKWARD: 5' GTATTACCCAGCTGGCGAAAGGGGATGTG 3' (M13FSP)

Insert Length: 1 Std Error: 0.00

Seq primer: 5'-CCGTCGGGAATTCCTCGGT-3' pSport3/86

High quality sequence stop: 629.

Location/Qualifiers

1. .629

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMG498E05107"

/tissue_type="whole embryo"

FEATURES

source

ORIGIN

/dev stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMP498)"
 /note="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGCGTCG-3' adapters (Gibco BRL)."

Alignment Scores:

Pred. No.:	5.07e-12	Length:	629
Score:	188.00	Matches:	53
Percent Similarity:	43.33%	Conservative:	25
Best Local Similarity:	29.44%	Mismatches:	88
Query Match:	15.49%	Indels:	14
DB:	4	Gaps:	4

US-10-006-922A-12 (1-225) x BI384961 (1-629).

Qy 3 SerSerlyshenValIleLysGluPheMetArgPheLysValArgMetGlu----- 19
 Db 60 TCATCCAAAGACAAATCGCGCAATCATGCTCTCCGACAAACCTCGACTTCGACATC 119

Qy 20 ---GlyThrValAenGlyHisGluPheGluLeuGlyGluGlyArgProTyr 38
 Db 120 TATGGCTCCATCAACGGTATGGAGTTGCATGTTGGCGGTGGAGTGGCAACCCAAAG 179

Qy 39 GluGlyHisAenThrValLysLeuLysValThrLysGly-----GlyProLeu 54
 Db 180 GACGATCGCTGACGTAACATGAGTCTACCAAGGTGCCCTGCGCGGTTCCTCTCTG 239

Qy 55 ProPheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLys 74
 Db 240 -----CTGGTGGCGCGCATCTGGGTAGCGCCACTATCATGACTACCTACC 284

Qy 75 HisProAlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGlu 94
 Db 285 TTCCTGACGCTCGCTCGCCTTTCCAGCGAGCTGTGAACACGGCGGTATCAATGCAT 344

Qy 95 ArgValMetAenPheGluAspGlyValValThrValThrGlnAspSerSerLeuGln 114
 Db 345 CGCTCTTCAACTTCGAGGAGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 404

Qy 115 AspGlyCysPheIleTyrLysValLysPheIleGlyValAenPheProSerAspGlyPro 134
 Db 405 GCGCGCAAGATCCATGAGAGTTTCATCTGTGGGCGAGCTGTTCCCGGACGATAGTCCG 464

Qy 135 ValMetGlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyr---ProArg 153
 Db 465 GTGATGACCAACGCGCTGACCGCTTTTGGACAGGAGCGTGGCCAGCTCATGTGCGTCC 524

Qy 154 AspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyr 173
 Db 525 GATGACAAGCTTGCGGAGTCTGTGGAGTGGACCTACCGCACTAGCAGCGCGCGCTAC 584

RESULT 11
 BI382442 707 bp mRNA linear EST 26-AUG-2003
 LOCUS BFLG2.000619 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMP498) Branchiostoma floridae cDNA clone MPMP498C08131 5', mRNA sequence.
 ACCESSION BI382442
 VERSION BI382442.1 GI:30917513
 KEYWORDS EST.
 SOURCE Branchiostoma floridae (Florida lancelet)
 ORGANISM Branchiostoma floridae
 Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
 REFERENCE 1 (bases 1 to 707)
 AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M. and Lehrach, H.

TITLE New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes

JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)
 MEDLINE 22683279
 PUBMED 12799346

COMMENT Contact: Panopoulou G
 Laboratory 145, dept. Lehrach
 Max-Planck-Institut fuer Molekulare Genetik
 Ihnestr.63-73, D-14195 Berlin, Germany
 Tel: +49 30 8413 1235
 Fax: +49 30 8413 1128
 Email: panopoulou@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
 Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers
 FORWARD: 5' CCCAGGCTTACACTTATGTTCCGGCTCG 3' (M13RSP)
 BACKWARD: 5' GGTATTACCCAGCTGGCGAAGAGGGGTGTG 3' (M13FSP)
 Insert Length: 1 Std Error: 0.00
 Seq primer: 5'-CCGTCGGAATTCCTCGGT-3' pSport3/86
 High quality sequence stop: 707.

FEATURES

source
 1..707
 /organism="Branchiostoma floridae"
 /mol_type="mRNA"
 /db_xref="taxon:7739"
 /clone="MPMP498C08131"
 /tissue_type="whole embryo"
 /dev_stage="5-6 hrs (gastrula stage)"
 /lab_host="E.coli, XLI blue"
 /clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMP498)"
 /note="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTTCTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGCGTCG-3' adapters (Gibco BRL)."

ORIGIN

Alignment Scores:

Pred. No.:	7.9e-12	Length:	707
Score:	187.00	Matches:	54
Percent Similarity:	42.78%	Conservative:	29
Best Local Similarity:	27.84%	Mismatches:	97
Query Match:	15.40%	Indels:	14
DB:	4	Gaps:	4

US-10-006-922A-12 (1-225) x BI382442 (1-707)

Qy 3 SerSerlyshenValIleLysGluPheMetArgPheLysValArgMetGlu----- 19
 Db 89 TCATCCAAAGACAAATCGCGCAACATCATGCTCTCCGACAAACCTCGACTTCGACATC 148

Qy 20 ---GlyThrValAenGlyHisGluPheGluLeuGlyGluGlyArgProTyr 38
 Db 149 TATGGCTCCATCAACGGTATGGAGTTGCATGTTGGCGGTGGAGTGGCAACCCAAAG 208

Qy 39 GluGlyHisAenThrValLysLeuLysValThrLysGly-----GlyProLeu 54
 Db 209 GACGATCGCTGACGTAACATGAGTCTACCAAGGTGCCCTGCGCGGTTCCTCTCTG 268

Qy 55 ProPheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLys 74
 Db -----CTGGTGGCGCGCATCTGGGTAGCGCCACTATCATGACTACCTACC 284

Db 269 -----CTGGTGGCGCCGATCTAGGTACGGCCACTACCTACCTACCC 313
Qy 75 HisProAlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGlu 94
Db 314 TTCCCTGATGTCGCTCGCTTTCCAGGCATCCGTGAACACGGCGGTATCAAAATGCAT 373
Qy 95 ArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeuGln 114
Db 374 CGCTCTTCAACTTTGAGGACGGTGGCTGCTGACTGCCACTACACTACTCTCTACAGC 433
Qy 115 AspGlyCysPheLysLysValLysValLysPheLysValLysPheProSerAspGlyPro 134
Db 434 GCGGCAAGATCCAGGAGAGATTTTCATCTGTGGGAGCTGTCTCCCGCAGATAGTCCG 493
Qy 135 ValMetGlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyr---ProArg 153
Db 494 GTGATGACCAACGGCGTGCAGCGGTCTGGACAGGAGCGTGCCCAAGCTGATGTCGGTCTCC 553
Qy 154 AspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyr 173
Db 554 GATGACAAGCTTGCAGAGTTGCTGGACTGGACCTACCTCACCAGCAGCGGGCGGTAC 613
Qy 174 LeuValGluPheLysSerIleTyrMetAlaLysLysProVal 187
Db 614 CGTGCCACGGTGCAGACCAACTTCACCTTCGCANAGCCCATC 655

RESULT 12

LOCUS BI380055 689 bp mRNA linear EST 26-AUG-2003
DEFINITION BFLG1_001686 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498N2369 5', mRNA sequence.

ACCESSION BI380055 GI:30915180

VERSION EST.

KEYWORDS Branchiostoma floridae (Florida lancelet)

SOURCE Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.

ORGANISM 1 (bases 1 to 689)

REFERENCE Panopoulou G

AUTHORS Laboratory 145, dept.Lehrach

TITLE Max-Planck-Institut fuer Molekulare Genetik

Innestr. 63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterized by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each

cluster is an indicator of the frequency of a transcript in the

analysed library. The cluster size as well the coordinates of the

rest of the clones assigned to the same fingerprint cluster as the

clone from which the above EST is generated is available at the

amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>

Clones and filters are distributed via the Resource Center/Primary

Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCGAGCTTTACACTTTATGCTTCGGCTCG 3' (M13RSP)

BACKWARD: 5' GTATTACCCGCTGGCGAAGGGGATGTG 3' (M13FSP)

Insert Length: 1200 Std Error: 0.00

Seq primer: 5'-CCGGTCCGGAATCCCGGT-3' pSport3/86

High quality sequence stop: 689.

FEATURES
source

1. 689
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPGp498N2369"
/issue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMPGp498)"
/notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTCTAGATCGGCGCGGCC (T)15-3' and a SalI 5'-TCGACCCACCGCTCCG-3' adapters (Gibco BRL)."

ORIGIN

Alignment Scores: 1.16e-11 Length: 689
Pred. No.: 185.50 Matches: 53
Score: 44.86% Conservative: 30
Percent Similarity: 28.65% Mismatches: 87
Best Local Similarity: 15.28% Indels: 15
Query Match: 4 Gaps: 5
DB:

US-10-006-922A-12 (1-225) x BI380055 (1-689)

Qy 16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluLeuGluGlyGluGly 35
Db 97 ATCCACTTTCACGGCTCCATCAACGCCACGAGTTCGACATGTTGGGGAGGAAAGGC 156
Qy 36 ArgProTyrGluGlyHisThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
Db 157 GACCCGAACGCCGGCTCGTGTGTGACACAGCAAGAAATCCACCAAGGGTGC---CTGAAG 213
Qy 56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
Db 214 TTCTCTCCCTACTTGTATGATCCCCACCTCGGGTACGGGTACTACGAGTACCTCCCTAC 273
Qy 76 ProAlaAspIleProAspTyrLysLysLeuSerPheProGlu-----GlyPheLysTrp 93
Db 274 CCG---GACGGACCTTCGCGCTTCCAGGCTCCATGTTGGAAGGATCGGGGTATGCCGTC 330
Qy 94 GluArgValMetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLeu 113
Db 331 TACCGCGTGTTCGAATTTGAAGACGAGGCAAGCTGACTGACGAGTTTAAGTACTCTCTAC 390
Qy 114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly 133
Db 391 GAGGGTTCCCATATCAAGCCGACATGAGGCTGATGGGAAGCGGTTTCCCTGACGACGGC 450
Qy 134 ProValMetGlnLysLysThrMetGlyTrpGluAlaSer---ThrGluArgLeuTyrPro 152
Db 451 CCAGTCATGACACCGCAGATTGTGACACGAGCGCTGCGTGTCCAGAGAGAGGTATCTTT 510
Qy 153 ArgAspGlyValLeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHis 172
Db 511 AACAGAACACCATCTGTGGACAGCTTCGACTGGAGATACAACTCGAGATGGGAAGCGC 570
Qy 173 TyrLeuValGlu-----PhyLysSerIleTyrMet 182
Db 571 TACAGGGCCGAGTGTGAGCCACTAAATCTTCGACAGCCCTTTTTCAGCCGATCTCATG 630
Qy 183 AlaLysProVal 187
Db 631 AAGAGCAGCGCGGTC 645

RESULT 13

BI387406

LOCUS

DEFINITION

BI387406 571 bp mRNA linear EST 26-AUG-2003
BFL26_002109 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMPGp531) Branchiostoma floridae cDNA clone MPMPGp531M1877 5', mRNA

```

sequence.
ACCESSION BI387406
VERSION BI387406.1 GI:30922264
EST.
KEYWORDS
SOURCE
ORGANISM Branchiostoma floridae (Florida lancelet)
          Branchiostoma floridae
          Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
          Branchiostoma
          1 (bases 1 to 571)
REFERENCE
AUTHORS Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
          Herwig, R., Vingron, M. and Lehrach, H.
TITLE New evidence for genome-wide duplications at the origin of
          vertebrates using an amphioxus gene set and completed animal
          genomes
JOURNAL Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE 22683279
PUBMED 12799346
COMMENT Contact: Panopoulou G
          Laboratory 145, dept. Lehrach
          Max-Planck-Institut fuer Molekulare Genetik
          Innestr.63-73, D-14195 Berlin, Germany
          Tel: +49 30 8413 1235
          Fax: +49 30 8413 1128
          Email: panopoul@molgen.mpg.de
          The library was characterised by oligonucleotide fingerprinting
          (ONFP) to reduce sequencing redundancy. According to the ONFP
          procedure, clones giving the same hybridisation pattern with a
          battery of 200 8mer oligonucleotides are grouped into clusters. One
          clone per cluster is selected for sequencing. The size of each
          cluster is an indicator of the frequency of a transcript in the
          analysed library. The cluster size as well the coordinates of the
          rest of the clones assigned to the same fingerprint cluster as the
          clone from which the above EST is generated is available at the
          amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
          Clones and filters are distributed via the Resource Center/Primary
          Database of the German Genome Project (http://www.rzp.de)
          PCR Primers
          FORWARD: 5' CCCAGGCTTACACTTATGCTCCGGCTCG 3' (M13RSP)
          BACKWARD: 5' GCATATACGCCAGTCGCGAAAGGGGATGTG 3' (M13FSP)
          Insert Length: 1 Std Error: 0.00
          Seq primer: 5'-CGGTCCGGAATTCGGGT-3' pSPORT3/86
          High quality sequence stop: 571.
FEATURES
          source
          Location/Qualifiers
            1..571
              /organism="Branchiostoma floridae"
              /mol_type="mRNA"
              /db_xref="taxon:7739"
              /clone="MPMGp531M1877"
              /tissue_type="whole embryo"
              /dev_stages="26 hrs (neurula stage)"
              /lab_host="E.coli, Xli blue"
              /clone_lib="Amphioxus 26hr cDNA library (Name convention:
              BFL26 or MPMGP531)"
              /note="Vector: pSPORT1 (Gibco BRL); Site 1: SalI, KpnI,
              EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT
              primed and directionally cloned in pSPORT1 vector using a
              NotI (5'-pGACTAGTTCTAGTCGCGAGCGGCC (T)15-3' and a
              SalI 5'-TCGACCCCGCTCGG-3' adapters (Gibco BRL)."
          Alignment Scores:
            Pred. No.: 1.18e-11 Length: 571
            Score: 184.50 Matches: 50
            Percent Similarity: 45.51% Conservative: 26
            Best Local Similarity: 29.94% Mismatches: 86
            Query Match: 15.20% Indels: 5
            DB: 4 Gaps: 4
          US-10-006-922A-12 (1-225) x BI387406 (1-571)
Qy 16 ValargMetGluGlyThrValAsnGlyHisGluPheGluLeuGluGlyGluGly 35
  ::: ::::::::::::::::::::|
Db 77 ATCCACCTTCACGGGTCCATCACAGCCACCGAGTTCGACATGTTGGTGGGGGAAAGGC 136

```

```

36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyProLeuPro 55
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
137 GACCCGAACCGCGCTCGCTGGTGACACACGCGAAATCCACCAAGGGTGCC--CTGAAG 193
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 TTCTCTCCCTACTGTATGATGCCCCACCTCGGGTACGGGTACTACCACTACCTCCCTAC 253
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
76 ProAlaAspIleProAspTyrLysLysLeuSerPheProGlu-----GlyPheLysTrp 93
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
254 CCG---GACGACCCCTCGCTTTCCAGACCTCCATGTTGGAAGGATCGGGGTATGCAGTC 310
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 GluArgValMetAsnPheGluAspGlyValValThrValThrGlnAspSerSerLeu 113
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 TACCGCGTGTTCGACTTTGAAGACGGAGCAAGCTGTCTACCGAGTTTAAGTACTCTCTAC 370
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly 133
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 GAGGTTCCCATATCAAGCGCGACATGAGCTGATGGAGAGCGGTTTCCCTGACGACGGC 430
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
134 ProValMetGlnLysLysThrMetGlyTrpGluAlaSerThrGlu---ArgLeuTyrPro 152
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 CCACTCATGACCCAGCAGATTGTGACACGAGCGCTGCGTCCATGAAGACGTATCTT 490
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
153 ArgAspGlyValLysLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHis 172
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
491 AACGACAACCACTCGTGGACAGCTTCGACTGGAGTTTCAACCTGCACNATGGGAAGCGC 550
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 TyrLeuValGluPheLysSer 179
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
551 TACAGGCGCGAGTGTCTGAGC 571
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 14
BO551001 550 bp mRNA linear EST 20-JUN-2002
LOCUS HA006G01-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
DEFINITION HA006G01-5, mRNA sequence.
ACCESSION BO551001
VERSION BO551001.1 GI:21451887
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 550)
AUTHORS VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G.,
          Martin, P.R., Stagg, C.A., Bassey, U., Alba, K., Hamatani, T.,
          Kargul, G.J., Luo, A.G., Kelso, J., Hide, W. and Ko, M.S.H.,
          Assembly, verification, and initial annotation of NIA 7.4K mouse
          cDNA clone set
          Genome Res. 12 (12), 1999-2003 (2002)
JOURNAL 22354164
MEDLINE 12466305
PUBMED
COMMENT Other_ESTs: HA006G01-3
          Contact: Yong Qian
          Laboratory of Genetics
          National Institute on Aging/National Institutes of Health
          333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
          Email: cdna@nlgsun.grc.nia.nih.gov
          This clone set has been freely distributed to the community. Please
          visit http://nlgsun.grc.nia.nih.gov/cDNA/NIA\_7\_4k.html for details.
          Plate: HA006 row: G column: 01
          Seq primer: -21M13 Reverse
          High quality sequence stop: 550
          POLYA=No.
          Location/Qualifiers
            1..550
              /organism="Mus musculus"
              /mol_type="mRNA"
              /strain="C57BL/6"
              /db_xref="niaEST:HA006G01-5"
              /db_xref="taxon:10090"

```

```

/clones="H4006G01"
/sex="mixed"
/dev stage="mixed"
/lab_hosts="DH10B"
/clone_lib="NIA Mouse 7.4K cDNA Clone Set"
/notes="vector: pSPORT1; Site_1: Sall; Site_2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3.41e-11 Length: 550
Score: 180.50 Matches: 38
Percent Similarity: 50.82% Conservative: 24
Best Local Similarity: 31.15% Mismatches: 57
Query Match: 14.87% Indels: 3
DB: 5 Gaps: 2

```

US-10-006-922A-12 (1-225) x BQ551001 (1-550)

```

Qy 16 ValArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGly 35
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 186 GTCAGCTGGACGGCGAGCGTAACGGCCACAGTTCAGCGTGTCCGCGAGGGCGAGGC 245
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 36 ArgProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuPro 55
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 246 GATGCCACCTACGCGAAGCTGACCTGAAGTTCATCTGCACACC---GGCAAGCTGCC 302
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 56 PheAlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHis 75
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 303 GTGCCCTGGCCACCCCTCGTACACCCCTGACCTACCGCGTGCAGTGTTCAGCGGTAC 362
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 76 ProAlaAspIlePro-----AspTyrLysLysLeuSerPheProGluGlyPheLysTrp 93
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 363 CCCGACCACTGAAGCAGCAGCAGCTTCTCAAGTCCGCCATGCCCAAGGCTACGTCAG 422
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 94 GluArgValMetAsnPheGluAspGlyValValThrValThrGlnAspSerSerLeu 113
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 423 GAGCGCACCATCTTCTTCAAGGACGACGGCACTACAAGACCCGCGCGAGGTGAAGTTC 482
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 114 GlnAspGlyCysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGly 133
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 483 GAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGCGC 542
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 134 ProVal 135
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 543 AACATC 548
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

RESULT 15

```

BI381780
LOCUS BI381780
DEFINITION BFLG1.003411 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPWGP498) Branchiostoma floridae cDNA clone MPWGP498C047 5',
mRNA sequence.

```

ACCESSION

```

VERSION BI381780
KEYWORDS BI381780.1 GI:30916875
SOURCE EST.

```

ORGANISM

```

Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

```

REFERENCE

```

1 (bases 1 to 671)
Panopoulou, G., Herwig, S., Groth, D., Krause, A., Poustka, A. J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes

```

Genome Res. 13 (6A), 1056-1066 (2003)

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Panopoulou G

Laboratory 145, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

lhnestr.63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>. Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)

FEATURES

source

```

1..671
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPWGP498C047"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name
convention: BFLG or MPWGP498)"
/notes="Vector: pSPORT1; Site_1: Sall, XpnI, EcoRI (5');
Site_2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSPORT1 vector using a NotI
(5'-pGACTAGTTCATGATCGGAGCGGCGCC (T)15-3' and a Sall 5'-
TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."
```

ORIGIN

```

Alignment Scores:
Pred. No.: 2.1e-10 Length: 671
Score: 175.00 Matches: 48
Percent Similarity: 46.78% Conservative: 32
Best Local Similarity: 28.07% Mismatches: 87
Query Match: 14.42% Indels: 4
DB: 4 Gaps: 3

```

US-10-006-922A-12 (1-225) x BI381780 (1-671)

```

Qy 20 GlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGlu 39
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 87 GGCTCGTCTAATGCCGCGGAGTTCGACCTGTGTGGAGGGGGAAGGCAACCCGAACGAT 146
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 40 GlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAsp 59
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 147 GGAACGCTGGAGACCAGTGTGAATCCACCCGCGGCGCC---CTGCCCTCTCCCGCTG 203
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 60 IleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIle 79
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 204 TTGATCGGACCCCAACCTCGGGGTACGGTTTCTACAGTACCTCTCCCTCGCGGCGCC 263
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 80 ProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPhe 99
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 264 TCACCTCTCCAAACCCGCATCAGGAGGAGGTACAGTTCAAGTTGTTCAGATTT 323
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 100 GluAspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIle 119
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 324 GAAGACGGGGAGTGTCTGAATCCAACTCCCGCTACACCTACGAGGGCGCGCAAGATCAAA 383
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

Qy 120 TyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLys 139
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 384 GGGAGTTCAGCTGATGATGGTTCAGTTTCCCTGCGCGGGGCGCTGTGTATGTCTCCGCGGA 443
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :

```

```
Qy 140 ThrMetGlyTrpGluAlaSerThrGluArgLeuTyr---ProArgAspGlyValLeuLys 158
Db 444 CTGACCACTGGACAGGAGCGTGGCCAAACTGCAGTCTCGGACGACCGCACCATCACC 503
Qy 159 GlyGluIleHisLysAlaLeuLysLeuLysAspGlyGly-----HisTyrLeuValGlu 176
Db 504 GGCACCAACAACCTGGAGCTTCTGCACCACCGATGGCGATGGGAAGCGCCACCGCGGAT 563
Qy 177 PheLysSerIleTyrMetAlaLysLysProVal 187
Db 564 GTGCAGACGAACTACACCTTCGCCAAGCGGCTC 596
```

Search completed: July 1, 2005, 20:51:19
Job time : 3346 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 1, 2005, 16:33:33 ; Search time 555 Seconds
(without alignments)
2399.895 Million cell updates/sec

Title: US-10-006-922a-12

Perfect score: 1214

Sequence: 1 MRSSKNVKEFMFKVMEG.....EDYIVEQYERTGRHHLFL 225

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xih
-Q=/cgn2_1/USPTO_spool/US10006922/runat_30062005_110616_19077/app_query.fasta_1.391
-DB=N Geneseq_16Dec04 -OPMT=fastap -SUFFIX=mg -MINMATCH=0.1 -LOOPECL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10006922 @CGN_1_1_470 @runat_30062005_110616_19077 -NCPV=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq_16Dec04:
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1214	100.0	678	AAH47654	AAH47654 Discosoma
2	1214	100.0	678	AAAD13053	AAAD13053 Discosoma
3	1214	100.0	678	AAAD11142	AAAD11142 Discosoma
4	1214	100.0	678	ABA95905	ABA95905 Yeast opt
5	1214	100.0	678	ABA95922	ABA95922 Yeast opt

6	1214	100.0	678	ABA95921	ABA95921 Yeast opt
7	1214	100.0	678	ABA95920	ABA95920 Yeast opt
8	1214	100.0	678	AAH46278	AAH46278 Discosoma
9	1214	100.0	678	AAAD28207	AAAD28207 Discosoma
10	1214	100.0	678	ADC24127	ADC24127 Discosoma
11	1214	100.0	678	ADF70404	ADF70404 Discosoma
12	1214	100.0	678	ADL46204	ADL46204 Discosoma
13	1214	100.0	678	ADN33978	ADN33978 Wild-type
14	1214	100.0	678	ADL36420	ADL36420 Discosoma
15	1214	100.0	678	ADM97768	ADM97768 D sp red
16	1214	100.0	695	AAA48743	AAA48743 Humanised
17	1214	100.0	859	AAAL47952	AAAL47952 Discosoma
18	1214	100.0	859	AAAD53432	AAAD53432 Discosoma
19	1214	100.0	859	AAAD61969	AAAD61969 Discosoma
20	1214	100.0	3311	ACA62995	ACA62995 Plasmid D
21	1210	99.7	681	ADC24134	ADC24134 Discosoma
22	1210	99.7	681	ADL46205	ADL46205 Human cod
23	1210	99.7	681	ADL46225	ADL46225 Human cod
24	1210	99.7	723	ABZ22476	ABZ22476 Mammalian
25	1210	99.7	1638	ADL18131	ADL18131 REP:PS(NI
26	1210	99.7	1647	ADL18155	ADL18155 RFP:PS(HI
27	1210	99.7	2721	ADS75466	ADS75466 Fibrohexa
28	1210	99.7	2772	ADS75468	ADS75468 Fibrohexa
29	1210	99.7	4692	AAAL47954	AAAL47954 Modified
30	1210	99.7	4692	ACC44640	ACC44640 Vector pD
31	1210	99.7	5436	AAAD10003	AAAD10003 Plasmid p
32	1210	99.7	6893	ADE24111	ADE24111 Proviral
33	1210	99.7	7910	AAAD09979	AAAD09979 pBit (dHSP
34	1210	99.7	9320	ABSS56664	ABSS56664 Plasmid p
35	1210	99.7	9658	ADL04104	ADL04104 Vector pT
36	1210	99.7	9678	ADL04103	ADL04103 Vector pT
37	1210	99.7	10283	ADL04102	ADL04102 Vector pT
38	1209	99.6	681	ABA95906	ABA95906 Yeast opt
39	1209	99.6	681	ABA95907	ABA95907 Yeast opt
40	1207	99.4	678	AAAD28208	AAAD28208 Discosoma
41	1205	99.3	681	AAH47656	AAH47656 Anthozoan
42	1203	99.1	7508	ADR30814	ADR30814 Zebrafish
43	1197	98.6	6985	ACA55359	ACA55359 Transform
44	1196	98.5	666	ABL61142	ABL61142 Red fluor
45	1196	98.5	711	ADO78067	ADO78067 Corallimo

ALIGNMENTS

RESULT 1

AAH47654

ID AAH47654 standard; cDNA; 678 BP.

AC AAH47654;

XX 30-NOV-2001 (first entry)

DT Discosoma sp. red drFP583 protein coding sequence.

DE Fluorescent protein; Anthozoan; fluorescence; marker; FRET; drFP583; ss.

XX Discosoma sp.

OS WO200162919-A1.

XX 30-AUG-2001.

PD 13-FEB-2001; 2001WO-US004625.

XX 23-FEB-2000; 2000US-0184732P.

XX (AURO-) AUROBA BIOSCIENCES CORP.

XX Nelson D, Zamaira E, Tsien R;

XX WPI; 2001-557704/62.

XX Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise

functional red fluorescent proteins, and the encoding nucleic acids, with key mutations for improving the proteins function.

Disclosure; Page 83; 90pp; English.

The invention provides a nucleic acid encoding functional red fluorescent CC protein (II) that differs from the sequence of an Anthozoan red fluorescent protein by at least one amino acid substitution, and with different fluorescent properties. The red fluorescent protein of the invention can be expressed by standard recombinant methodology. (II) are used a fluorescent markers and FRET partners. It is used for identifying protein-protein interactions. (II) is also suitable for multiplexed fluorescent analysis and FRET-based applications using existing Aquorea fluorescent proteins. (II) has improved brightness, reduced spectral cross talk, and is rapidly and efficiently expressed in mammalian cells. The key mutations in the encoding nucleic acids provide improved folding, brightness, and create (II) with sharper, more defined excitation and emission peaks when expressed in mammalian cells. The present sequence CC represents the coding sequence of a *Discosoma* sp. drFP583 protein, an anthozoan fluorescent protein

Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-006-922A-12 (1-225) x AAH47654 (1-678)

```
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCCTCCAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTTCGCATCGAAGGA 60
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCATATGGCAGCAGTTTGAATAGAGCGGAGGAGGAGGAGGCCATACGAGGC 120
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLysProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGAGCCCTTGCCATTTCGCTGGGATATT 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTACCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCTGCCGACATACCA 240
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAGCTGTCATTTCTCGAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCGTTACTGTACCCAGGATTCAGTTTGCAAGGATGGCTGTTTCATCTAC 360
Qy 121 LysValLysPheIleGlyValAsnPheProSerHepGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCATTTGGCGTGAACCTTCCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTGATGGCGTGTGAAGGAGAG 480
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCATAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAGAGCCCTGTGACGTACCGGGTACTACTATGTTGTACTCCAACTGGAT 600
```

```
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAAAGCCACAAACAGACTATACAATCGTTGACAGCTATGAAGAACCGAGGACGC 660
```

```
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTTCCTT 675
```

RESULT 2

AAID13053 standard; cDNA; 678 BP.

AAID13053;

16-OCT-2001 (first entry)

Discosoma sp. "red" anthozoa fluorescent protein, drFP583 cDNA.

Protein destabilisation; linker moiety; reporter moiety; disease model;
linear multimerised domain; -NH-ubiquitin protein endoprotease;
transgenic animal; transgenic plant; disease resistance;
anthozoa fluorescent protein; natural fluorescent protein; ss.

Discosoma sp.

WO200157242-A2.

09-AUG-2001.

02-FEB-2001; 2001WO-US003791.

04-FEB-2000; 2000US-00498098.

(AURO-) AUROBA BIOSCIENCES CORP.

Stack JH, Whitney M, Cubitt AB, Pollok BA;

WPI; 2001-488890/53.

Destabilizing proteins in living cells, by coupling a target protein to linear multimerized destabilization domain non-cleavable by -NH-ubiquitin protein endoproteases, comprising two copies of the domain.

Disclosure; Page 110; 171pp; English.

The present invention relates to a method for destabilising a target protein in a cell. The method comprises a linker moiety which operatively couples a target protein (a reporter moiety) to a linear multimerised destabilising domain, which is non-cleavable by a -NH-ubiquitin protein endoproteases. The method is useful for detecting an activity such as protease, protein kinase or phosphoprotein phosphatase activity and is also useful for identifying novel assays for a wide range of post-translational activities, such as proteolysis, phosphorylation, dephosphorylation, glycosylation, methylation, sulfation, prenylation, disulfide bond formation and ADP-ribosylation within cells. The recombinant DNA molecule of the invention is useful for creating transgenic animals useful as disease models and transgenic plants with improved disease resistance or other favourable traits. The present sequence is *Discosoma* sp. "red" anthozoa fluorescent protein, drFP483 cDNA which is a natural fluorescent protein used as a reporter moiety in the exemplification of the invention

Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0

US-10-006-922A-12 (1-225) x AAD13053 (1-678)

```
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAAGATGTTATCAAGAGTTTCATGAGTTTAAGGTTTCGATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCCAGAGTTTGAATAGAGGCGAAGAGGAGGAGGCCATACGAAGGC 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGCTTAAGTTAAGTAACCAAGGGGACCTTTGCCATTTGCTGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCGAAGGAGGAGGAGGCCATACGAAGGC 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATATAAAGCTGTCAATTTCTGAAGGATTTAAATGGGAAGGGGTTCATGAACCTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheLeGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCATTTGGCGTGAACCTTTCTTCGATGGACCTGTGTATGCAAAAGAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCATGAGCGTTGTATCTCTCGTATGGCGTGTGTTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAAATTCAAAGTATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAAGCCGTGTGCAGTACCAGGGTACTACTATGTGTGACTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAAACAGCCACACGAGCACTACAACTCGTTGAGCAGTATGAAGAAGCCGAGGAGCG 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTCTCTT 675
```

RESULT 3

```
AAD11142
ID AAD11142 standard; DNA; 678 BP.
XX
AC AAD11142;
XX
XX
DT 24-SEP-2001 (first entry)
XX
DE Discosoma species Anthozoa fluorescent protein, drFP583 encoding DNA.
XX
KW Transmembrane potential; biological membrane; fluorescent ion; detection;
KW test chemical screening; Anthozoa fluorescent protein; FP;
KW transgenic organism; drFP583 protein; ds.
XX
OS Discosoma sp.
XX
PN W0200142211-A2.
XX
PD 14-JUN-2001.
XX
PF 12-DEC-2000; 2000WO-US033739.
XX
PR 13-DEC-1999; 99US-00459956.
```

```
XX
PA (REGC ) UNIV CALIFORNIA.
```

```
XX
PI Tsien RY, Gonzalez JE;
```

```
XX
PI WPI; 2001-457276/49.
```

```
XX
PT Determining electrical potential across a membrane in biological systems,
PT comprises introducing two reagents, exposing the membrane to light and
PT measuring the energy transfer.
```

```
XX
PS Disclosure; Page 150; 154pp; English.
```

```
XX
CC The patent discloses optical methods and compositions for determining
CC transmembrane potential across biological membranes in living cells. The
CC method of determining the electrical potential across a membrane in a
CC biological system comprises introduction of two reagents, a first reagent
CC comprising a hydrophobic fluorescent ion capable of redistributing from a
CC first face of the membrane to a second face of the membrane in response
CC to membrane potential change and a second reagent that label the first
CC face or the second face of the membrane, which comprises a chromophore
CC capable of undergoing energy transfer by either donating or accepting
CC excited state energy to the fluorescent ion. The membrane is then exposed
CC to excitation light and the energy transfer between the reagents is
CC measured and related to the membrane potential. The method is useful for
CC detecting changes in membrane potential in subcellular organelle
CC membranes in biological systems. The method is used for screening of test
CC chemicals for activity to modulate the activity of target ion channel.
CC The invention also provides a transgenic organism comprising a first
CC reagent that comprises a charged hydrophobic fluorescent molecule and a
CC second reagent comprising a bioluminescent or naturally fluorescent
CC protein. The present sequence is Discosoma species (red) DNA encoding an
CC Anthozoa fluorescent protein (FP), drFP583
```

```
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

Alignment Scores:

```
Pred. No.: 2,02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 4 Gaps: 0
```

```
US-10-006-922A-12 (1-225) x AAD11142 (1-678)
```

```
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAAGATGTTATCAAGGAGTTTCATGAGTTTAAGGTTTCGATGGAAGGA 60
```

```
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCCAGAGTTTGAATAGAGGCGAAGAGGAGGAGGCCATACGAAGGC 120
```

```
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGACCTTTGCGATTTGCTGGATATT 180
```

```
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCGAAGGATATATGTCAAGCACCTTCGCGCATACCA 240
```

```
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATATAAAGCTGTCAATTTCTGAAGGATTTAAATGGGAAGGGGTTCATGAACCTTTGAA 300
```

```
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360
```

```
Qy 121 LysValLysPheLeGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCATTTGGCGTGAACCTTTCTCTCGATGGAGCCTGTTATGCAAAAGAGACA 420
```

QY 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Db 421 ATGGGCTGGAGAGCAGCATGAGCGTTTGTATCCTCGTATGCGGTGTTGAAGAGAG 480
 QY 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 Db 481 ATTCAAGAGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAGATT 540
 QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValLysSerLysLeuAsp 200
 Db 541 TACATGGCAAGAGCGCTGTCAGCTACCGAGGTACTATGTGTGATCTCAAACTGGAT 600
 QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Db 601 ATACACAGCCCAACGAACTATACAACTGTTGAGCAGTATGAAGACCGAGGACGC 660
 QY 221 HisHisLeuPheLeu 225
 Db 661 CACCATCTGTTCTT 675

RESULT 4

ABA95905
 ID ABA95905 standard; cDNA; 678 BP.

XX AC ABA95905;

DT 29-MAY-2002 (first entry)

XX Yeast optimised RFP encoding cDNA SEQ ID NO 1.

DE Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;

KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;

KW Escherichia coli; green fluorescent protein; biotechnology; gene; ss.

XX Anthozoa.

XX Key Location/Qualifiers

FH 1--678

FT CDS /*tag= a

FT /product= "yeast optimised Red Fluorescent Protein"

FT DE20001395-U1.

XX 15-MAR-2001.

XX 27-JAN-2000; 2000DE-02001395.

XX 27-JAN-2000; 2000DE-02001395.

XX (GPCB-) GPC BIOTECH AG.

DR WPI; 2002-228394/29.

DR P-PSDB; ABB08834.

XX New DNA encoding red fluorescent protein, useful as marker in
 PT biotechnology, has sequence optimized for expression in eukaryotes,
 PT especially yeast or plants.

XX Claim 1; Fig 1; 19pp; German.

XX The invention relates to DNA (I) containing either sequence ABA95905 or
 CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
 CC (YRFP). (I) are used to express red fluorescent protein (RFP) in
 CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
 CC plants, especially dicotyledonous plants including Nicotiana tabacum or
 CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
 CC especially Escherichia coli. RFP is useful in the same way as green
 CC fluorescent protein but is more generally applicable in modern
 CC biotechnology. (I) are optimised for expression in yeast and so generate
 CC RFP at higher levels with stronger fluorescence and thus lowers the
 CC detection limit and gives a better signal-to-noise ratio

SQ Sequence 678 BP; 198 A; 147 C; 159 G; 174 T; 0 U; 0 Other;

Alignment Scores: 2.02e-140 Length: 678

Pred. No.: 1214.00 Matches: 225

Score: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x ABA95905 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

Db 1 ATGAGATCTTCTAAGAACGTCATCAAGGAATTCATGAGATTCAGAGTTAGAATGGAAGGT 60

QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40

Db 61 ACTGTTAAGCGTCCAGCATTCGAAATCGAAGGTGAAGGTCAAGGTAGACCATACGAAGGT 120

QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTTPAspIle 60

Db 121 CACAACACTGTCAAGTTGAAGGTACTAAGGGTGGTCCATTGCCATTTCGTTGGGACATC 180

QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Db 181 TTGTCTCCACNATTCCAATACGGTCTTAAGGTCTACGTCAAGCACCAGCTGACATTCCTCA 240

QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrPheLysValMetAsnPheGlu 100

Db 241 GACTACAGAAGTTGTCTTCCAGAAAGGTTTCAAGTGGGAAAGAGTCATGAATTCGAA 300

QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Db 301 GACGGTGGTGTGTGTACTGTACTCAAGACTCTCTTTCGAAAGACGGTGTTCATCTAC 360

QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Db 361 AAGGTCAAGTTCATCGTGTCAACTTCCCATCTGACGGTCCAGTCATGCAAAAGAGACT 420

QY 141 MetGlyTyrPheLysLeuSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

Db 421 ATGGGTTGGGAGAGCTTCTACCGAACGTTTGTACCAAGAGACGGTGTCTTGAAGGGTGAA 480

QY 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

Db 481 ATCCACAAGCGCTTGAAGTTGAAGGACGGTGGTCACTACTTGGTCCGAATTCAGATCTATC 540

QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

Db 541 TACATGGCTAAGAAAGCCAGTCCCAATTCACAGGTCTACTACGTCTAAGTTGGAC 600

QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220

Db 601 ATCACCTCTCACAAACGAAGACTACACTATCGTCCGAACAATACGAGCTACTGAAGGTAGA 660

QY 221 HisHisLeuPheLeu 225

Db 661 CACCATCTGTTCTT 675

RESULT 5

ABA95922

ID ABA95922 standard; DNA; 678 BP.

XX AC ABA95922;

DT 29-MAY-2002 (first entry)

DE Yeast optimised RFP related DNA SEQ ID NO 19.

XX Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
 KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
 KW Escherichia coli; green fluorescent protein; biotechnology; gene; ss.

CC especially Escherichia coli. RFP is useful in the same way as green
CC fluorescent protein but is more generally applicable in modern
CC biotechnology. (I) are optimised for expression in yeast and so generate
CC RFP at higher levels with stronger fluorescence and thus lowers the
CC detection limit and gives a better signal-to-noise ratio. The present
CC sequence is that of a polynucleotide encoding the yeast optimised RFP,
CC useful to the invention
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x ABA95921 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGTCTTCCAGAAATGTTATCAGAGTTTCAAGGTTTAAAGTTTCGATGAAGGA 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
DB 61 ACGTCAATGGCCACGAGTTTGAATAGAGGCGAAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTyrAspIle 60
DB 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGAGCTTTGCCATTTTGGTGGATATT 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
DB 181 TTGTCCACCAATTTCTAGTATGGAAGCAAGGTATATGTCAAGCACTTCCGACATACCA 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnPheGlu 100
DB 241 GACTATAAAAGCTGTCTATTTCTGTAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 300
QY 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
DB 301 GACGGTGGCGTCTGTACTTAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
DB 361 AAGTCAAGTTCATTCGCGTGAACCTTCTTCCATGACCTGTTATGCAAGAGACCA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
DB 421 ATGGCTGGGAAGCCAGCACTAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
DB 481 ATTCATAGGCTCTGAAGCTGAAGAGCGGTGTCATTTACCTAGTTGAATTCAAAGTATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
DB 541 TACATGCAAGAAGCCTGTGCAGCTACCGGGTACTACTATGTTGACTCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
DB 601 ATRACAGCCACACGAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGGAGGACGC 660
QY 221 HisHisLeuPheLeu 225
DB 661 CACCATCTGTTCTTT 675

RESULT 7

ABA95920
ID ABA95920 standard; RNA; 678 BP.

XX
AC ABA95920;

XX 29-MAY-2002 (first entry)
XX
DE Yeast optimised RFP encoding RNA SEQ ID NO 16.
XX
XX Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
KW Escherichia coli; green fluorescent protein; biotechnology; ss.
XX
XX Anthozoa.
XX
XX DE20001395-U1.
XX
XX 15-MAR-2001.
XX
XX 27-JAN-2000; 2000DE-02001395.
XX
XX 27-JAN-2000; 2000DE-02001395.
XX
XX (GPCB-) GPC BIOTECH AG.
XX
XX WPI; 2002-228394/29.
XX
PT New DNA encoding red fluorescent protein, useful as marker in
PT biotechnology, has sequence optimized for expression in eukaryotes,
PT especially yeast or plants.
XX
PS Disclosure; Page 13; 19pp; German.
XX
CC The invention relates to DNA (I) containing either sequence ABA95905 or
CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
CC (YRFP). (I) are used to express red fluorescent protein (RFP) in
CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
CC plants, especially dicotyledonous plants including Nicotiana tabacum or
CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
CC especially Escherichia coli. RFP is useful in the same way as green
CC fluorescent protein but is more generally applicable in modern
CC biotechnology. (I) are optimised for expression in yeast and so generate
CC RFP at higher levels with stronger fluorescence and thus lowers the
CC detection limit and gives a better signal-to-noise ratio. The present
CC sequence is that of an RNA sequence corresponding to the yeast optimised
CC RFP encoding cDNA given as SEQ ID NO 1 (ABA95905)
XX
SQ Sequence 678 BP; 198 A; 147 C; 159 G; 0 T; 174 U; 0 Other;

Alignment Scores:
Pred. No.: 2.02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x ABA95920 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 AUGAGAUCUUCUAAAGAACGUCACUAGGAUUCUAGAGAUUCAAAGUUAAGGAAGGU 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
DB 61 ACUGUUAACGUCACGACGAAUUCGAAUUCGAAAGGUGGAGGAGGAGGAGGAGGAGG 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTyrAspIle 60
DB 121 CACAACUGUCAAGUUAAGGAGGUGUUAUUAAGGGUGGUGGUGGUGGUGGUGGUGG 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
DB 181 UUGUCCACAAUCCAAUACGUCUAGGUCUACGUCAGGAGGAGGAGGAGGAGGAGGAG 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTyrGluArgValMetAsnPheGlu 100

RESULT 9
 AAD28207
 ID AAD28207 standard; DNA; 678 BP.
 XX
 AC AAD28207;
 XX
 DT 22-APR-2002 (first entry)
 XX
 DE Discosoma sp. humanised wild-type Anthozoa protein drFP583 DNA.
 XX
 KW Fluorescent timer protein; protein movement; translocation; trafficking;
 KW promoter activity; gene expression; transgenic plant; gene modification;
 KW protein age; anthozoa protein; drFP583; ds.
 XX
 OS Discosoma sp.
 XX

Key Location/Qualifiers
 CDS 1..678
 FT /*tag= a
 FT /product= "Humanised wild-type Anthozoa protein drFP583"
 XX
 PN WO200196373-A2.
 XX
 PD 20-DEC-2001.
 XX
 XX 13-JUN-2001; 2001WO-US019097.
 PF
 XX 14-JUN-2000; 2000US-0211607P.
 PR
 XX (CLON-) CLONTECH LAB INC.
 PA
 XX Fradkov AF, Tersikh A;
 PI
 XX WPI; 2002-154595/20.
 DR
 XX P-PSDB; AAE17540.
 XX

New fluorescent timer proteins comprising an emission spectrum that changes over time from a first wavelength to a second wavelength, useful for monitoring intracellular protein movement, translocation, trafficking or stability.

Example 1; Fig 1; 89pp; English.

The invention relates to a fluorescent timer protein having an emission spectrum that changes over time after synthesis from a first wavelength to a second wavelength. The fluorescent timer proteins are useful in monitoring the activity of a promoter, determining the age of a protein, identifying an agent that modulates the activity of a promoter and in enriching a population of cells comprising a fluorescent timer protein. The fluorescent timer proteins are also useful for assessing gene expression during development of a multicellular organism or during cellular differentiation, in response to a drug or other inducer of promoter activity, as a reporter to serve as a read-out of promoter activity, monitoring intracellular protein movement or translocation, protein trafficking, or protein stability, to investigate temporal aspects of the activity of a regulatory element, for determining cell fate during development and organ remodelling, in spatial and temporal visualisation of newly synthesised proteins and accumulated proteins, and in distinguishing between newly formed and pre-existing structures, e.g. membrane junctions and extracellular matrix components. The fluorescent timer proteins may further be used to investigate where photobleaching techniques are employed, as detectable labels, as selectable markers, as biosensors in prokaryotic and eukaryotic cells, in protease cleavage assays, and as second messenger detectors. The nucleic acids can be used to generate transgenic, non-human plants or animals or site-specific gene modifications in cell lines. The present sequence is a DNA encoding Discosoma sp. humanised wild-type Anthozoa protein drFP583 used for generating fluorescent proteins

SQ Sequence 678 BP; 146 A; 223 C; 203 G; 106 T; 0 U; 0 Other;
 , Alignment Scores:

Pred. No.: 2,02e-140 Length: 678
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0

US-10-006-922A-12 (1-225) x AAD28207 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 |||||
 DB 1 ATCGCGTCTCTCAAGAACGTCATCAAGGAGTTTCATCGCTTCAAGTGCGCATGGAGGC 60
 |||||
 QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
 |||||
 DB 61 ACCGTGAACGGCCACGAGTTCGAGATCGAGGGGAGGGGCGGCCCTTACGAGGGC 120
 |||||
 QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 |||||
 DB 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGGGGGCCCCCTCGCCTTGGCTGGGACATC 180
 |||||
 QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 |||||
 DB 181 CTGTCCCCCAGTTCCAGTACGGCTCCAAGGTGTACGTGAAGCACCCTCCGACATCCCC 240
 |||||
 QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 |||||
 DB 241 GACTACAGNAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAATTCGAG 300
 |||||
 QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 |||||
 DB 301 GACGGCGGGTGGTGACCGCTGACCCAGGACTCTCTCCCTGCAGGACGGCTTTCATCTAC 360
 |||||
 QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 |||||
 DB 361 AAGGTGAAGTTTCATCGCGGTGAACCTTCCCTCCGACGGCCCCGTGATGCAGAGAAGACC 420
 |||||
 QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 |||||
 DB 421 ATGGGTGGGAGGCGCTCCACCGAGCGCTGTACCCCCGGCGAGCGCGTGCTGAAGGGCGAG 480
 |||||
 QY 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 |||||
 DB 481 ATCCACAAAGGCCCTGAAGCTGAAGGACGGCGGCACTACTCTGGTGGAGTTCAAGTCCATC 540
 |||||
 QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 |||||
 DB 541 TACATGGCCAAAGAGCCCGTGCAGCTGCCCGGCTACTACTACGTGGACTCCAAGCTGGAC 600
 |||||
 QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 |||||
 DB 601 ATCACCCTCCCAACAGGAGNCTACACCATCGTGGAGCAGTACGAGCGCACCGAGGGCGCG 660
 |||||
 QY 221 HisHisLeuPheLeu 225
 |||||
 DB 661 CACCACCTGTTCCTG 675

RESULT 10
 ADC24127
 ID ADC24127 standard; DNA; 678 BP.
 XX
 AC ADC24127;
 XX
 DT 18-DEC-2003 (first entry)
 XX
 DE Discosoma wild-type red fluorescent protein DNA #1.
 XX
 KW Discosoma red fluorescent protein; DsRed; AB interface; AC interface;
 KW fluorescence protein variant; transcription induction detection;
 KW fluorescence energy resonance transfer; FRET; protein kinase;
 KW protein phosphatase; ion indicator; ds.
 XX
 OS Discosoma.

XX US2003059835-A1.
 PN 27-MAR-2003.
 XX 10-APR-2002; 2002US-00121258.
 XX 26-FEB-2001; 2001US-00794308.
 PR 24-MAY-2001; 2001US-00866538.
 XX (TSIE//) TSIE R Y.
 PA (CAMP//) CAMPBELL R E.
 XX Tsien RY, Campbell RE;
 PI WPI; 2003-743764/70.
 XX P-PSDB; ADC24126.
 DR Novel polynucleotide sequence encoding Discosoma red fluorescent protein
 PT variant having a reduced propensity to oligomerize, useful for detecting
 PT transcriptional activity.
 XX Example 1; SEQ ID NO 2; 67pp; English.
 CC The invention describes a polynucleotide sequence (I) encoding a
 CC Discosoma red fluorescent protein (DsRed) variant having a reduced
 CC propensity to oligomerize, comprising amino acid substitutions at the AB
 CC and/or AC interfaces of the wild-type DsRed sequence (S1) comprising 225
 CC amino acids, given in the specification, where the substitutions result
 CC in reduced propensity of the DsRed variant to form tetramers. (I) is
 CC useful for detecting transcriptional activity by providing a host cells
 CC containing a vector which comprises (I) operatively linked to an
 CC expression control sequence, and an unit to assay the variant fluorescent
 CC protein fluorescence, and assaying fluorescence of the variant
 CC fluorescent protein produced by (VII), where variant fluorescent protein
 CC fluorescence is indicative of transcriptional activity. A polynucleotide
 CC encoding a fusion protein is useful for the analysis of in vivo
 CC localisation or trafficking of a polypeptide of interest. A polypeptide
 CC marker is useful as markers to identify the location and amount of a
 CC target protein produced, where the target protein is fused to the marker,
 CC as a complement to or alternative for the green fluorescent protein or
 CC its spectral variant, for detecting induction of transcriptions, in
 CC applications involving fluorescence energy resonance transfer (FRET),
 CC which detects events as the function of the movement of fluorescent
 CC donors and acceptors towards or away from each other, for making
 CC fluorescent sensors for protein kinase and phosphatase activities or
 CC indicators for ions and molecules such as Ca²⁺, Zn²⁺, for identifying the
 CC presence of a molecule in a sample, for identifying a specific
 CC interaction of a first and second molecule, for determining whether a
 CC sample contains an enzyme or for determining the pH of the sample. (I) is
 CC useful for identifying a region or condition that regulates the activity
 CC of an expression control sequence. This sequence encodes Discosoma wild-
 CC type red fluorescent protein.
 XX Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
 SQ
 Alignment Scores:
 Pred. No.: 2, 02e-140 Length: 678
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservatives: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 10 Gaps: 0
 US-10-006-922A-12 (1-225) x ADC24127 (1-678)
 QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 DB 1 ATCAGGTCTTCCCAAGATGTTATCAGAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
 QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
 DB 61 ACGGTCAATGGGCACGAGTTTGAATAGAGGGCGAAGGAGGGGGGAGGCCATACGAAGGC 120

QY 41 HisAsnThrValLysLeuLysValThrLysGlyProLeuProPheAlaTrpAspIle 60
 DB 121 CACAAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGGCATTTTGGGATATT 180
 QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 DB 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCGACATACCA 240
 QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 DB 241 GACTATAAAAAGCTGTCTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
 QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 DB 301 GACGGTGGCGTCGTTACTGTAAACCCAGGATTCAGATTTCAGGATGGCTGCTTTTCATCTAC 360
 QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 DB 361 AAGTCAAGTTCAATTCGCGTGAACCTTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
 QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLysLysGlyGlu 160
 DB 421 ATGGCGTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGAAAGGAGAG 480
 QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 DB 481 ATTCAATAAGGCTCTGAAGCTGAAAGACGGTGGTCAATTACCTAGTTGAATTCAAAGATT 540
 QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 DB 541 TACATGGCAAAAGAGCGCTGGCAGCTACCAAGGTACTACTATGTTGACTCCAAACTGGAT 600
 QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 DB 601 ATAAACAGCCACACCAAGACACTATACAACTGTTGAGCAGTATGAAGAACCGAGGACGC 660
 QY 221 HisHisLeuPheLeu 225
 DB 661 CACCATCTGTTCTCTT 675
 RESULT 11
 ADF70404
 ID ADF70404 standard; cDNA; 678 BP.
 XX AC ADF70404;
 DT 12-FEB-2004 (first entry)
 XX Discosoma wild-type GFP variant cDNA SeqID27.
 DE ligand; orphan receptor protein; fusion protein; fluorescent protein; GFP; GFP-1; wild-type GFP;
 KW cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
 KW GFPuv; Enhanced GFP; EGFP; gene; ss.
 XX Discosoma sp.
 OS WO2003071272-A1.
 PN 28-AUG-2003.
 PD 21-FEB-2003; 2003WO-JP001901.
 PF 22-FEB-2002; 2002JP-00045728.
 PR 23-JUL-2002; 2002JP-00213949.
 PR 11-OCT-2002; 2002JP-00298237.
 XX (TAKE) TAKEDA CHEM IND LTD.
 XX Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
 PI WPI; 2003-697654/66.
 XX P-PSDB; ADF70403.
 DR

XX Transformation of cells with a fusion protein of an orphan receptor
PT protein with a fluorescent protein useful for identification of ligands
PT to the orphan receptor.
XX
PS Disclosure; SEQ ID NO 27; 594pp; Japanese.
XX
CC This invention relates to a novel method of identifying ligands to an
CC orphan receptor protein which comprises transforming cells with DNA
CC encoding a fusion protein of the orphan receptor with a fluorescent
CC protein, so that the fusion protein is expressed in the cells (or cell
CC membranes isolated from them) and contacting the cells with the potential
CC ligand to be tested. A suitable fluorescent protein for incorporation in
CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,
CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
CC identification of ligands binding to an orphan receptor protein.
XX
SQ Sequence 678 BP; 204 A; 129 C; 179 G; 166 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0
US-10-006-922A-12 (1-225) x ADF70404 (1-678)
QY 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGCTTTCCAAAGATTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
DB 61 ACGGTCAATGGCGACGAGTTTGAATAGAGGCGAAGGAGGAGGAGGAGCCATACGAAGGC 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLysProPheAlaTyrAspIle 60
DB 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGGTTGGATATT 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
DB 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysThrGluArgValMetAsnPheGlu 100
DB 241 GACTATAAAGAGCTGCTCATTTCTCGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
DB 301 GACGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheLeGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
DB 361 AAGGTCAAGTTTCATTGGCGTGAACTTTCTCCGATGGACCTGTGTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
DB 421 ATGGGCTGGGAAGCCACCTGACGCTGTGTATCCTCGTATGGCGGTGTGAAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
DB 481 ATTCAATAGGCTCTGAAGCTGGAAGACGGTGGTCACTTACCTAGTTGATTCATAAGATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
DB 541 TACATGGCAAGAAGCCTGTGCAGCTACCGAGGTACTACTATGTTGACTCCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
DB 601 ATAACAAGCCACACGAAGACTATACAATCGTTGAGCAGTATGAAGAACCGAGGACGC 660

QY 221 HisHisLeuPheLeu 225
DB 661 CACCATCTGTTCTT 675
RESULT 12
ADL46204
ID ADL46204 standard; DNA; 678 BP.
XX AC ADL46204;
XX 20-MAY-2004 (first entry)
XX Discosoma red fluorescent protein (DsRed) coding sequence.
DE ds: gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
XX oligomerization; tetramerization; immunoassay; hybridization assay.
XX Discosoma sp.
XX Key Location/Qualifiers
FT CDS 1..678
FT /*tag= a /product= "DsRed protein"
XX PN WO2003086446-A1.
XX PD 23-OCT-2003.
XX PF 09-APR-2003; 2003WO-US010879.
XX PR 10-APR-2002; 2002US-00121258.
XX PR 29-JUL-2002; 2002US-00209208.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Tsien RY, Campbell RE, Baird GS;
XX WPI; 2003-845265/78.
XX P-PSDB; ADL46203.
XX New monomeric and dimeric Anthozoan fluorescent protein variants with
XX reduced propensity to oligomerize, and encoding polynucleotides, useful
XX in molecular biology, e.g. in immunoassays or in tracking protein
XX movement in cells.
XX Disclosure; SEQ ID NO 2; 166pp; English.
XX The invention relates to a polynucleotide sequence encoding a Discosoma
XX red fluorescent protein (DsRed) variant having a reduced propensity to
XX oligomerize. The protein variant comprises one or more amino acid
XX substitutions at the AB and/or AC interface(s) of the wild-type DsRed
XX sequence, where the substitutions result in reduced propensity of the
XX DsRed variant to form tetramers and where the variant displays detectable
XX fluorescence of at least one red wavelength. The composition and methods
XX are useful in producing red fluorescent proteins having reduced
XX propensity for oligomerization, especially tetramerization. The protein
XX may be used in molecular biology and in other scientific applications,
XX such as in immunoassays or hybridization assays, or in tracking the
XX movement of proteins in cells. This sequence corresponds to the DsRed
XX coding sequence.
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 11 Gaps: 0
US-10-006-922A-12 (1-225) x ADL46204 (1-678)

```
QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGCTCTCCAAAGATTTATCATCAGAGTTTCATGAGGTTTAAAGTTTCGATCGAAGGA 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
DB 61 ACGGTCAATGGCAGCAGTTTGAATAGAAGGCGAAGGAGGAGGAGGCCATACGAAGGC 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
DB 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGATATT 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
DB 181 TTGTCCACCAATTTTCAGTATGAGCAAGATATATGTCAAGCACCTTCGCGACATACCA 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
DB 241 GACTATAAAAGCTGTCTATTTCTCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300
QY 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
DB 301 GACGGTGGCGTCTACTTAACCCAGGATTCAGATTTGCGAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
DB 361 AAGTCAAGTTCATTTGGCGGAACCTTTCTTCGATGGACCTGTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
DB 421 ATGGCTGGGAGCCAGCAGCTAGCGTTTGTATCTCTGATGGCGTGTGAAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
DB 481 ATTCAATAGGCTCTGAAGCTGAAAGACGGTGGTTCATTACCTAGTTGAATTCAAAGTATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
DB 541 TACATGGCAAGAAGCCCTGTGCAGCTACCGAGGTTACTATGTGTGACTCCAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
DB 601 ATACAGCCACACAGAGACTATACATTTGTTGAGCAGTATGAAAGACCGAGGAGCG 660
QY 221 HisHisLeuPheLeu 225
DB 661 CACCATCTGTTCCCTT 675
RESULT 13
ADN33978
ID ADN33978 standard; DNA; 678 BP.
XX
AC ADN33978;
XX
DT 01-JUL-2004 (first entry)
XX
DE Wild-type DARED encoding sequence.
XX
KW Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DARED;
KW ds.
XX
OS Discoboma sp.
XX
PN WO2003054158-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-US040539.
XX
PR 19-DEC-2001; 2001US-0341723P.
XX
PA (UYCH-) UNIV CHICAGO.
XX
```

```
PI Bevis B, Glick B;
XX WPI; 2003-569236/53.
DR P-PSDB; ADN33979.
XX
PT Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT mutant of a Cnidarian chromo- or fluorescent protein or its mutant,
PT useful for applications involving chromo- or fluorescent proteins.
XX
PS Claim 8; SEQ ID NO 1; 65pp; English.
XX
CC The present invention relates to nucleic acid that encodes a rapidly
CC maturing chromo or fluorescent mutant of a Cnidarian chromo- or
CC fluorescent protein or its mutant. The protein is useful in applications
CC involving nucleic acid encoding a chromo- or fluorescent protein and is
CC useful for producing a chromo and/or fluorescent protein which involves
CC growing the cell, whereby the protein is expressed, and isolating the
CC protein substantially free of other proteins. The protein is useful in
CC applications involving chromo- or fluorescent protein and is useful as
CC PCR primers, hybridization probes, etc. The expression cassettes are
CC useful for synthesizing related proteins. The chromoproteins are useful
CC as coloring agents which are capable of imparting color or pigment to a
CC particular composition of matter e.g. food compositions, pharmaceuticals,
CC cosmetics, living organisms, e.g., animals and plants. The chromoproteins
CC may also find use as labels in analyte detection assays e.g. assays for
CC biological analytes of interest and as selectable markers in recombinant
CC DNA applications, e.g. the production of transgenic cells and organisms.
CC The fluorescent proteins find use in a variety of different applications, as
CC e.g. in fluorescence resonance energy transfer (FRET) applications, as
CC biosensors in prokaryotic and eukaryotic cells, in applications involving
CC the automated screening of arrays of cells expressing fluorescent
CC reporting groups by using microscopic imaging and electronic analysis, as
CC second messenger detectors, and in fluorescence activated cell sorting
CC applications and as in vivo marker in animals. The fluorescent proteins
CC also find use in protease cleavage assays. The proteins can also be used
CC as assays to determine the phospholipid composition in biological
CC membranes and as a fluorescent timer. The present sequence represents the
CC wild-type DARED encoding sequence.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.: 2,02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
```

US-10-006-922A-12 (1-225) x ADN33978 (1-678)

```
QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGCTCTCCAAAGATTTATCATCAGGAGTTTCATGAGGTTTAAAGTTTCGATCGAAGGA 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
DB 61 ACGGTCAATGGCAGCAGTTTGAATAGAAGGCGAAGGAGGAGGAGGCCATACGAAGGC 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
DB 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGATATT 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
DB 181 TTGTCCACCAATTTTCAGTATGAGCAAGATATATGTCAAGCACCTTCGCGACATACCA 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
DB 241 GACTATAAAAGCTGTCTATTTCTCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300
QY 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
```

Db 301 GACGCTGGCGCTGTTACTGTAAACCCAGGATTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
QY 121 LysValIysPheIleGlyValAlaSerThrGluArgLeuTyrProArgAspGlyProValMetGlnLysLysThr 140
Db 361 AAGTCAAGTTCATTGGCGTGAACCTTTCTTCGATGGACCTGTTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTCATGGCGTGTGAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAAGCCTGTGACGCTACCAAGGCTACTATGTTGACTCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAAAGCCACCAAGCACTATACAACTCGTTGAGCAGTATGAAGAACCGAGGACGC 660
QY 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCTT 675

RESULT 14
ADI36420
ID ADI36420 standard; DNA; 678 BP.
XX
AC ADI36420;
XX
DT 22-APR-2004 (first entry)
XX
DE Discosoma sp. red fluorescent protein (RED) DNA.
XX
KW Fluorobody; binding ligand; green fluorescent protein; GFP;
KW target detection; gene; red fluorescent protein; ds; RED.
XX
OS Discosoma sp.
XX
FH Key Location/Qualifiers
FT CDS 1..678
FT /*tag= a
FT /product= "Discosoma sp. red fluorescent protein (RED) "
XX
PN US2003203355-A1.
XX
XX 30-OCT-2003.
XX
XX 24-APR-2002; 2002US-00132067.
XX
XX 24-APR-2002; 2002US-00132067.
XX
XX (LALA-) LOS ALAMOS NAT LAB.
XX (REGC) UNIV CALIFORNIA.
XX
XX Bradbury AM, Zeytun A, Waldo GS;
XX
XX WPI; 2004-154325/15.
XX P-PSDB; ADI36421.
XX
XX Novel binding ligand with intrinsic fluorescence and comprising
XX fluorescent protein having heterologous binding sites, useful for
XX detecting target molecule.
XX
XX Example 6; SEQ ID NO 3; 23pp; English.
XX
XX The invention relates to binding ligands (fluorobodies) with intrinsic
XX fluorescence, which comprises green fluorescent protein (GFP) having
XX heterologous binding sites. The binding ligand is useful for detecting
XX the target molecule and is efficiently detects the target molecule. The
XX present sequence is Discosoma sp. red fluorescent protein (RED) DNA used

CC in the exemplification of the invention.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2,02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0
US-10-006-922a-12 (1-225) x ADI36420 (1-678)
QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAAGTATGTTATCAAGGAGTTTATGAGGTTTAAAGTTTCCATGGGAAGA 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGTCAATGGGCACCGAGTTTGAATAGAAAGCGCAAGGAGGAGGAGGCCATACGAAGGC 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACATACCGTAAAGCTTAAGGTAAACCAAGGGGGGACCTTTGCCATTTCCTGGGATATT 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCCACCACAATTCAGTATGGAACAAGGTATATGTCAAGCACCCTGCCGACATACCA 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAGCTGTCATTTCCTGAAGGATTTAAATGGAAAGGTCATGAACCTTTGAA 300
QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLysGlnAspGlyCysPheIleTyr 120
Db 301 GACGCTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheIleGlyValAlaSerPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGTCAAGTTCATTGGCGTGAACCTTTCCTTCGATGGACCTGTTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATGCGCTGTTGAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAAGGCTCTGAAGCTGAAAGACGGTGGTGCATTACCTAGTTGAATTCAAAAGTATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAAGCCTGTGACGCTACCAAGGCTACTATGTTGACTCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAAAGCCACCAAGCACTATACAACTCGTTGAGCAGTATGAAGAACCGAGGACGC 660
QY 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCTT 675

RESULT 15
ADM97768
ID ADM97768 standard; DNA; 678 BP.
XX
XX ADM97768;
XX
XX 01-JUL-2004 (first entry)
XX
XX D sp red fluorescent protein coding sequence SEQ ID NO: 21.
XX ds; gene; enzyme; sensor cell; fluorescent protein;

KW signal transduction detection system; promoter; targeting sequence;
KW targeted drug.

OS Discosoma sp.

FH Key Location/Qualifiers

FT 1. .678

FT /tag= a

FT /product= "fluorescent protein"

XX

XX WO2004031415-A2.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PT Developing a sensor cell, useful in determining the activity of a target
PT gene and in developing therapeutic drugs, comprises providing cells
PT comprising a signal transduction detection system and introducing DNA
PT construct into cells.

PS Disclosure; Page 167-168; 234pp; English.

XX The present invention relates to a method of developing a sensor cell,
CC for determining the activity of a target gene in the cell, which
CC comprises providing a homogeneous population of cells, where each of the
CC cells comprises a signal transduction detection system and introducing
CC into the population of cells an isolated DNA construct comprising a
CC promoter operatively linked to a targeting sequence. The method is useful
CC in developing a sensor cell for determining the activity of a target gene
CC in the cell. The sensor cell and the methods are useful in developing new
CC and therapeutic drugs directed to the targets. The present sequence is a
CC coding sequence shown in the exemplification of the invention.

XX SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2.02e-140 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 12 Gaps: 0

US-10-006-922A-12 (1-225) x ADM97768 (1-678)

QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
DB 1 ATGAGGCTTTCCAGAAATGTTATCAAGAGGTTTCATGAGGTTTAAGGTTTCGATGGAAGGA 60
QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
DB 61 ACGGTCAATGGGCACGAGTTTGAATAGAGGCGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
DB 121 CACAAATCCGTAAGCTTAAGGTAAACAGGGGGGACCTTTGCAATTTGCTGGGATATT 180
QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
DB 181 TTGTCCACCAATTTCAAGTATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

Search completed: July 1, 2005, 18:42:03
Job time : 558 secs

Db 241 GACTATATAAAGCTGTCTATTTCTGAAGGATTTAAATGGGAAAGGTCATGAACTTTGAA 300
QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGGCTCGTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360
QY 121 LysValLysPheIleGlyValLeuPheProSerAspGlyProValMetGlnLysLeuThr 140
Db 361 AAGGTCAAGTTTCATTCGGGTGAACCTTCTCTCCGATGGACCTGTATGCAAAAGAGACA 420
QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTGATGCGTGTGAAAGGAGAG 480
QY 161 IleHisLysAlaLeuLysLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAATAGGCTCTGAAGCTGAAAGACGGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 540
QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAAAGAGCGCTGTGCAGCTACCAGGGGTACTACTATGTTGACTCCAAACTGGAT 600
QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAACCAAGCCACACCAAGCACTATACATCGTTGAGCAGTATGAAAGAACCGAGGAGCC 660
QY 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCTCT 675

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 1, 2005, 18:32:43 ; Search time 614 Seconds
(without alignments)
2298.439 Million cell updates/sec

Title: US-10-006-922a-12
Perfect score: 1214
Sequence: 1 MRSSKNVKEFMFKVMEG.....EDYIVEQYERTGRHHLFL 225

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US1006922/runat_30062005_110617_19138/app_query.fasta_1.391
-DB=Published Applications NA -QFMT=fastap -SURFIX=rnpb -MINMATCH=0.1
-LOOPCI=0 -LOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US1006922 @CGN 1.1 480 @runat_30062005_110617_19138
-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1214	100.0	678	9	US-09-967-772-6	Sequence 11, Appl
2	1214	100.0	678	13	US-10-006-922-11	Sequence 6, Appl
3	1214	100.0	678	14	US-10-081-864-7	Sequence 7, Appl
4	1214	100.0	678	14	US-10-121-258-2	Sequence 2, Appl
5	1214	100.0	678	16	US-10-315-920-1	Sequence 1, Appl
6	1214	100.0	678	17	US-10-132-067-3	Sequence 3, Appl
7	1214	100.0	678	17	US-10-335-517-6	Sequence 6, Appl
8	1214	100.0	678	17	US-10-334-288-6	Sequence 6, Appl
9	1214	100.0	678	19	US-10-311-030-5	Sequence 5, Appl
10	1214	100.0	678	21	US-10-656-029-21	Sequence 27, Appl
11	1214	100.0	678	21	US-10-505-486-27	Sequence 21, Appl
12	1214	100.0	859	9	US-09-999-745-66	Sequence 66, Appl
13	1214	100.0	859	10	US-09-866-538-11	Sequence 11, Appl
14	1214	100.0	859	10	US-09-794-308-11	Sequence 11, Appl
15	1214	100.0	859	10	US-09-865-291-11	Sequence 11, Appl
16	1214	100.0	859	19	US-10-433-640-12	Sequence 12, Appl
17	1214	100.0	859	20	US-10-885-988-11	Sequence 11, Appl
18	1214	100.0	859	21	US-10-857-622-11	Sequence 11, Appl
19	1214	100.0	3311	10	US-09-797-496B-3	Sequence 3, Appl
20	1210	99.7	681	13	US-10-006-922-35	Sequence 35, Appl
21	1210	99.7	681	14	US-10-121-258-3	Sequence 3, Appl
22	1210	99.7	681	14	US-10-121-258-23	Sequence 23, Appl
23	1210	99.7	681	19	US-10-311-030-8	Sequence 8, Appl
24	1210	99.7	713	19	US-10-311-030-11	Sequence 11, Appl
25	1210	99.7	723	14	US-10-152-296-1	Sequence 12, Appl
26	1210	99.7	723	15	US-10-739-656-1	Sequence 1, Appl
27	1210	99.7	723	15	US-10-214-932-51	Sequence 51, Appl
28	1210	99.7	1638	15	US-10-214-932-75	Sequence 75, Appl
29	1210	99.7	4692	15	US-10-161-403-29	Sequence 29, Appl
30	1210	99.7	4692	19	US-10-433-640-16	Sequence 16, Appl
31	1210	99.7	5436	21	US-10-169-050-46	Sequence 46, Appl
32	1210	99.7	6984	13	US-10-001-189-45	Sequence 45, Appl
33	1210	99.7	7910	21	US-10-169-050-20	Sequence 20, Appl
34	1210	99.7	9320	19	US-10-471-065-20	Sequence 20, Appl
35	1210	99.7	9658	19	US-10-609-019-4	Sequence 4, Appl
36	1210	99.7	9678	19	US-10-609-019-3	Sequence 3, Appl
37	1210	99.7	10263	19	US-10-315-920-3	Sequence 2, Appl
38	1203	99.1	7508	16	US-10-742-828-4	Sequence 3, Appl
39	1203	99.1	7508	19	US-10-006-922-37	Sequence 4, Appl
40	1198	98.7	681	13	US-10-332-733-22	Sequence 22, Appl
41	1196	98.5	686	19	US-10-314-936-1	Sequence 1, Appl
42	1196	98.5	711	19	US-10-314-936-3	Sequence 3, Appl
43	1196	98.5	711	23	US-11-021-014-1	Sequence 1, Appl
44	1196	98.5	711	23		
45	1196	98.5	711	23		

ALIGNMENTS

RESULT 1

US-09-967-772-6
; Sequence 6, Application US/09967772
; Patent No. US20020164577A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/09/967,772
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 08/765,860
; PRIOR FILING DATE: 1996-12-19

; PRIOR APPLICATION NUMBER: PCT/ US96/09652
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: US 08/481,977
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
US-09-967-772-6

Alignment Scores:
Pred. No.: 4,4e-143 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-10-006-922A-12 (1-225) x US-09-967-772-6 (1-678)

```
Qy 1 MetArgSerSerLysAsnValLleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCCTCCAAAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATGGAAGGA 60
Qy 21 ThrValAsnGlyHisGluPheGluLeuGluGluGluGluGluGluGluGluGluGluGlu 40
Db 61 ACGGTCATGGGACGAGTTTGAATAGAGGCGGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTtpAspIle 60
Db 121 CACAAATCCGTAAAGCTTAAGTAACCAAGGGGGGACCTTTCATTTTGGTATATT 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGCACCCCTGCCGACATACCA 240
Qy 81 AspTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAGCTGTCTATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCGTACTGTATGCCAGGATTCACGATTTGCGATGGCTGTTTCATCTAC 360
Qy 121 LysValLysPheLleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAAGTTTCATTTGGCGTGAACCTTTTCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCTGACGCTTTGATCCTCGATGGCGTGTGTTGAAGAGGAGAG 480
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCATAAGGCTCTGAAGCTGAAGAGCGGTGGTTCATTTACCTAGTTGAATTCAAAGATATT 540
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAGAGCCCTGTGTCAGCTACCGGGTACTACTATGTTGACTCTCAAACTGGAT 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATACAAAGCCACACGAAAGACTATACAACTCTTGAGCAGTATGAAGAACCGGAGGACGC 660
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCTCTT 675
```

RESULT 2

US-10-006-922-11

; Sequence 11, Application US/10006922

; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A.
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma species
US-10-006-922-11

Alignment Scores:
Pred. No.: 4,4e-143 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-006-922A-12 (1-225) x US-10-006-922-11 (1-678)

```
Qy 1 MetArgSerSerLysAsnValLleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCCTCCAAAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATGGAAGGA 60
Qy 21 ThrValAsnGlyHisGluPheGluLeuGluGluGluGluGluGluGluGluGluGluGlu 40
Db 61 ACGGTCATGGGACGAGTTTGAATAGAGGCGGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTtpAspIle 60
Db 121 CACAAATCCGTAAAGCTTAAGTAACCAAGGGGGGACCTTTCATTTTGGTATATT 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGCACCCCTGCCGACATACCA 240
Qy 81 AspTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAGCTGTCTATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 101 AspGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCGTACTGTATGCCAGGATTCAGTTTGGAGGATGGCTGTTTCATCTAC 360
Qy 121 LysValLysPheLleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAAGTTTCATTTGGCGTGAACCTTTTCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCTGACGCTTTGATCCTCGATGGCGTGTGTTGAAGAGGAGAG 480
```


Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAATAGGCTCTGAAGCTGAAGACGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 540
Qy 181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAGCCCTGTCAGCTACCGGCTACTACTATGTTGATCTCAAACTGGAT 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAAACAGCCCAACAGAGACTATACATCTGTGAGCAGTATGAAGAAACCGAGGGAGGC 660
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCCCTT 675

RESULT 3

US-10-081-864-7
; Sequence 7, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey
; APPLICANT: Lukyanov, Konstantin
; APPLICANT: Yanushevich, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; TYPE: DNA
; ORGANISM: Discosoma sp
US-10-081-864-7

Alignment Scores:

Pred. No.:	4,4e-143	Length:	678
Score:	1214.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-006-922A-12 (1-225) x US-10-081-864-7 (1-678)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGGCTCTCCAAAGAGCTCATCAAGGAGTTTCATCGCTTCAAGGTGCGCATGGAGGCG 60
Qy 21 ThrValAsnGlyHisGluPheGluLeuGlyGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACCGTGAACGCCACAGTTCGAGATCGAGGCGGAGGCGGAGGCGGCGGCTTACGAGGCG 120
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAACACCGTGAAGTGAAGTGAAGGCGGCGGCGGCGGCGGCTTTCGCTTGGGACATC 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 CTGTCCCCCAGTTCAGTACGGCTCCAAAGGTGTACGTGAAGACACCCCGCGGACATCCCC 240
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAGCTGTCTCTCCCCAGGGCTTCAAGTGGGAGCGCGGTGATGAATCTCGAG 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGCGCGCGTGGTGGTACCGTGAAGCCAGGACTCCTCCCTGCAGGACGGCTGCTTCATCTAC 360
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AGGTGAAGTTCATCGCGGTGAACCTTCCCTCCGACGCGGCCCGGTGATGCAAGAAAGACC 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGCTGGGAGGCTCCACCGAGCGCTGTACCCCGCGAGCGCGTGTGTAAGGGCGAG 480
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCACAAGGCCCTGAAGCTGAAGGACGCGGCGGCACCTACCTGGTGGATTCACAGTCCATC 540
Qy 181 TyrMetAlaLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAGCCCTGTCAGCTGCGGCTACTACCTGGGACTCCAGTGGAGTCCAGCTGAC 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACTCCCAACAGGAGTACACCATCTGTGGAGCAGTACGAGCGCACCAGCGGCGCGC 660
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACCTGTTCCCTG 675

RESULT 4

US-10-121-258-2
; Sequence 2, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(678)
; OTHER INFORMATION: wild-type DeRed
US-10-121-258-2

Alignment Scores:

Pred. No.:	4,4e-143	Length:	678
Score:	1214.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	14	Gaps:	0

US-10-006-922A-12 (1-225) x US-10-121-258-2 (1-678)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
Qy 21 ThrValAsnGlyHisGluPheGluLeuGlyGluGlyGluGlyArgProTyrGluGly 40

```
Db 61 ACGGTCAATGGGCACGAGCTTTGAAATAGAGCGCAAGGAGGAGGAGGCCATACGAAGGC 120
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAGCTTAAGAGTAACCAAGGGGGGACCTTTGCCATTTGGCTTGGATATT 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACACCAATTTCAATATGGGAAGCAAGATATATGTCAGACACCTTCCGCGACATACCA 240
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATATAAAGCTGTCTATTTCTGAAGCATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCAATGGGTGAACTTTTCCTTCCGATGGACCTGTTATGCCAAGAGACA 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGAAAGGAGAG 480
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGCGCAAGAGCCCTGTCACCTACCGGCTACTACTATGTTGATCTCCAACTGGAT 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAAACAGCCACAACGAAGACTATACAATCTTGCAGCAGTATGAAGAAACCGAGGAGCG 660
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCCCTT 675
```

RESULT 5

```
US-10-315-920-1
; Sequence 1, Application US/10315920
; Publication No. US20030175809A1
; GENERAL INFORMATION:
; APPLICANT: Fradkov, Arcady Fedorovich
; APPLICANT: Teresikh, Alexey
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS
; FILE REFERENCE: CLON-077CIP
; CURRENT APPLICATION NUMBER: US/10/315,920
; PRIOR FILING DATE: 2002-12-09
; PRIOR FILING DATE: 2000-06-14
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: variant of sequence from Discosoma sp.
US-10-315-920-1
```

Alignment Scores:

```
Pred. No.: 4,4e-143 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
```

```
Query Match: 100.00% Indels: 0
DB: 16 Gaps: 0
US-10-006-922a-12 (1-225) x US-10-315-920-1 (1-678)
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATCGGCTCTCCCAAGAACGTCATCAAGGAGTTTCATCGCTTCAAGGTGCGCATGGAGGC 60
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACCGTGAACCGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGCCGCCCTTACGAGGC 120
Qy 41 HisAsnThrValLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAACACCGTGAAGCTGAAGGTGACCAAGGGGGGGCCCCCTGCCCTTGGCTGGACATC 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 CTGTCCCCCAGTTCCAGTACGGCTCCAAAGGTGTAGTGAAGCACCCCGCCGACATCCCC 240
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTACAAGAAAGCTGTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACCTTCGAG 300
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGGGGGCGTGTGTGACCGTGACCCAGGACTCTCTCCCTGCAGGACGGCTGCTTCATCTAC 360
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTCAATCGGCTGAACTTCCCTCCGACGGCCCCCGTGTATGAGAGAGACC 420
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGGACGCGCGCTGCTGAAGGGCGAG 480
Qy 161 IleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCAAGAGCCCTGAAGCTGAAGGACGGGGGGCCACTACTCTGGTGAGTTCAAGTCCATC 540
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGCGCAAGAGCCCGCTGCAGCTGCCCGGCTACTACTACGTGACTCCAAGCTGGAC 600
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATCACCCTCCCAACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 660
Qy 221 HisHisLeuPheLeu 225
Db 661 CACCACCTGTTCTCTG 675
```

RESULT 6

```
US-10-132-067-3
; Sequence 3, Application US/10132067
; Publication No. US20030203355A1
; GENERAL INFORMATION:
; APPLICANT: Bradbury, Andrew
; APPLICANT: Zeytun, Ahmet
; APPLICANT: Waldo, Geoffrey
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic
; FILE REFERENCE: 021362-000600US
; CURRENT APPLICATION NUMBER: US/10/132,067
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
```

FEATURE:
 ; OTHER INFORMATION: red fluorescent protein (dsRED)
 ;
 ;
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(678)
 ; OTHER INFORMATION: dsRED
 US-10-132-067-3

Alignment Scores:
 Pred. No.: 4,4e-143 Length: 678
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

US-10-006-922A-12 (1-225) x US-10-132-067-3 (1-678)

```

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCCTCCAAAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCGACGAGTTTGAATAGACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGAGCCCTTTGCCATTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAAGCTGTCATTTCTGAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGTGGCGTCGTTACTGTAAACAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTTCATTTCCGATGAGACCTGTTTCCGATGAGACCTGTTTATGCAAAAGAGACA 420

Qy 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGCTGGGAAGCCAGCACCTGAGCGTTTGATCTCTGATGGCGTGTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCATAGGCTCTGAAGCTGAAGACGCTGATACCTACCTAGTTGAATTCAAAGATATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAAAGCCCTGTCAGCTACCAAGGCTACTACTGTTGATCTCCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220
Db 601 ATACACGCCAACAGACGATATACAAATCGTTGAGCAGATATGAAGAAGCCGAGGAGCGC 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCCCTT 675
  
```

RESULT 7

US-10-335-517-6
 ; Sequence 6, Application US/10335517
 ; Publication No. US20030207248A1
 ; GENERAL INFORMATION:
 ; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
 ; APPLICANT: TSIGEN, Roger

APPLICANT: GONZALEZ, Jesus
 ; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
 ; FILE REFERENCE: REGEN1290-5
 ; CURRENT APPLICATION NUMBER: US/10/335,517
 ; CURRENT FILING DATE: 2002-12-31
 ; PRIOR APPLICATION NUMBER: US/09/967,772
 ; PRIOR FILING DATE: 2001-09-28
 ; PRIOR APPLICATION NUMBER: US 09/459,956
 ; PRIOR FILING DATE: 1999-12-13
 ; PRIOR APPLICATION NUMBER: US 08/765,860
 ; PRIOR FILING DATE: 1996-12-19
 ; PRIOR APPLICATION NUMBER: PCT/ US96/09652
 ; PRIOR FILING DATE: 1996-06-06
 ; PRIOR APPLICATION NUMBER: US 08/481,977
 ; PRIOR FILING DATE: 1995-06-07
 ; NUMBER OF SEQ ID NOS: 22
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 678
 ; TYPE: DNA
 ; ORGANISM: Discosoma sp "red"
 US-10-335-517-6

Alignment Scores:
 Pred. No.: 4,4e-143 Length: 678
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 17 Gaps: 0

US-10-006-922A-12 (1-225) x US-10-335-517-6 (1-678)

```

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCCTCCAAAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCGACGAGTTTGAATAGACGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGAGCCCTTTGCCATTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAAGCTGTCATTTCTGAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGTGGCGTCGTTACTGTAAACAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTTCATTTCCGATGAGACCTGTTTCCGATGAGACCTGTTTATGCAAAAGAGACA 420

Qy 141 MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGCTGGGAAGCCAGCACCTGAGCGTTTGATCTCTGATGGCGTGTGAAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCATAGGCTCTGAAGCTGAAGACGCTGATACCTACCTAGTTGAATTCAAAGATATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGGCAAGAAAGCCCTGTCAGCTACCAAGGCTACTACTGTTGATCTCCAAACTGGAT 600
  
```

```
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGATGTAAGAAACCGAGGAGCG 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCTT 675

RESULT 8
US-10-334-288-6
; Sequence 6, Application US/10334288
; Publication No. US20040002123A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TS'EN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/10/334,288
; CURRENT FILING DATE: 2002-12-31
; PRIOR APPLICATION NUMBER: US/09/967,772
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/459,956
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 08/765,860
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT/ US96/09652
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: US 08/481,977
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
US-10-334-288-6

Alignment Scores:
Pred. No.: 4,4e-143 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 17 Gaps: 0

US-10-006-922A-12 (1-225) x US-10-334-288-6 (1-678)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAGATGTTTATCAAGGAGTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCCACGAGTTTGAATAAGAGCGCAAGGAGGAGGAGGCCATACGAAGGC 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACATACCGTAAAGCTTAAGTAACCAAGGGGAGCCCTTGCCATTTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTCAATATGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATATAAAGCTGTCTATTTCCCTGAAGGATTTAAATGGGAAGGGTTCATGAACCTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCGTCGTACTGTAAACCCAGGATTCAGATTTCCAGATGGCTGTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
```

```
Db 361 AAGGTCAAAGTTCAATGGCGTGAACCTTCTTCGATGGACCTGTATATGCAAAAGAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGGCGTGTGTAAGAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCATTACCTAGTTGAATTCAAAGATATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGCAAAAGAGCGCTGTGCAGCTACCAAGGGTACTACTATCTTGACTCAAACTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGATGTAAGAAACCGAGGAGCG 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCTT 675

RESULT 9
US-10-311-030-5
; Sequence 5, Application US/10311030
; Publication No. US20040171107A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamiatra, Elize
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/184,732
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
US-10-311-030-5

Alignment Scores:
Pred. No.: 4,4e-143 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 19 Gaps: 0

US-10-006-922A-12 (1-225) x US-10-311-030-5 (1-678)

Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAGATGTTTATCAAGGAGTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCCACGAGTTTGAATAAGAGCGCAAGGAGGAGGAGGCCATACGAAGGC 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACATACCGTAAAGCTTAAGTAACCAAGGGGAGCCCTTGCCATTTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTCAATATGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240
```

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
|
Db 241 GACTATATAAAGCTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
|
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
|
Db 301 GACGGTGGCGTCTGTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
|
Qy 121 LysValLysPheLleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
|
Db 361 AAGGTCAGTTCATTTGGCGTGAACCTTCCTCCGATGGACCTGTATATGCAAAAGAGACA 420
|
Qy 141 MetGlyTTPGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
|
Db 421 ATGGCTGGGAAGCAGCAGCTACCTGTTGATCTCTGTATGGCGGTGTTGAAAGGAGAG 480
|
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
|
Db 481 ATTCATAAGGCTCTGAAGCTGAAGACGGTGGTCTATACCTAGTTGAATTCANAAGTATT 540
|
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
|
Db 541 TACATGGCAAGAAGCCTGTGCAGCTACCGGGTACTACTATGTGTGACTCCAAACTGGAT 600
|
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
|
Db 601 ATACAAGCCACACGAAGACTATACAATCGTTGAGCAGATATGAAGAACCGAGGGACGC 660
|
Qy 221 HisHisLeuPheLeu 225
|
Db 661 CACCATCTGTCTCT 675
|

RESULT 10

US-10-656-029-21
; Sequence 21, Application US/10656029
; Publication No. US20050003367A1
; GENERAL INFORMATION:
; APPLICANT: VERTEX PHARMACEUTICALS INC.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RAPID DEVELOPMENT OF
; FILE REFERENCE: VPI/02-143W02
; CURRENT APPLICATION NUMBER: US/10/656,029
; PRIOR FILING DATE: 2003-09-05
; PRIOR APPLICATION NUMBER: 60/408,297
; PRIOR FILING DATE: 2002-09-05
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(678)
; OTHER INFORMATION: fluorescent protein
US-10-656-029-21

Alignment Scores:

Pred. No.:	4.4e-143	Length:	678
Score:	1214.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-10-006-922A-12 (1-225) x US-10-656-029-21 (1-678)

Qy 1 MetArgSerLysAsnValLleLysGluPheMetArgPheLysValArgMetGluGly 20
|
Db 1 ATGAGGCTTCCAAAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
|
Qy 21 ThrValAsnGlyHisGluPheGluLleGluGlyGluGlyArgProTyrGluGly 40
|

Db 61 ACGTCAATGGGCACGAGTTTGAATAGAACGCGAAGGAGGAGGCGCCATACGAAGGC 120
|
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTTPAspIle 60
|
Db 121 CACAATAACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
|
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
|
Db 181 TTGTACCACCAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTTCGCGACATACCA 240
|
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
|
Db 241 GACTATATAAAGCTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
|
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
|
Db 301 GACGGTGGCGTCTGTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
|
Qy 121 LysValLysPheLleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
|
Db 361 AAGGTCAGTTCATTTGGCGTGAACCTTCCTTCGATGGACCTGTATGCAAAAGAGACA 420
|
Qy 141 MetGlyTTPGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
|
Db 421 ATGGCTGGGAAGCAGCAGCTACCTGTTGATCTCTGTATGGCGGTGTTGAAAGGAGAG 480
|
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
|
Db 481 ATTCATAAGGCTCTGAAGCTGAAGACGGTGGTCTATACCTAGTTGAATTCANAAGTATT 540
|
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
|
Db 541 TACATGGCAAGAAGCCTGTGCAGCTACCGGGTACTACTATGTGTGACTCCAAACTGGAT 600
|
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
|
Db 601 ATACAAGCCACACGAAGACTATACAATCGTTGAGCAGATATGAAGAACCGAGGGACGC 660
|

RESULT 11

US-10-505-486-27
; Sequence 27, Application US/10505486
; Publication No. US20050118639A1
; GENERAL INFORMATION:
; APPLICANT: Takeda Chemical Industries, Ltd.
; TITLE OF INVENTION: Determination of a ligand
; FILE REFERENCE: P03-0006PCT
; CURRENT APPLICATION NUMBER: US/10/505,486
; CURRENT FILING DATE: 2004-08-20
; PRIOR APPLICATION NUMBER: JP 2002-45728
; PRIOR FILING DATE: 2002-02-22
; PRIOR APPLICATION NUMBER: JP 2002-213949
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: JP 2002-298237
; PRIOR FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 27
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp.
US-10-505-486-27

Alignment Scores:

Pred. No.:	4.4e-143	Length:	678
Score:	1214.00	Matches:	225
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-10-006-922A-12 (1-225) x US-10-505-486-27 (1-678)

Qy	1	MetArgSerSerIysAsnValIleYsGluPheMetArgPheLysValArgMetGluGly	20
Db	1	ATGAGGTCTTCCAAGAATGTATCAAGAGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA	60
Qy	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluArgProTyrGluGly	40
Db	61	ACGGTCAATGGGCACGAGTTTGAAATAGAAAGCCGAAAGAGGGGAGGCCATACGAAGGC	120
Qy	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Db	121	CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGGACCTTTGCCATTGTGCTGGGATATT	180
Qy	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValIleHisProAlaAspIlePro	80
Db	181	TTGTGCACCAAAATTTCAGTATGGGAAGCAAGGTATATGTCAAGACCCCTGCCGACATACCA	240
Qy	81	AspTyrLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Db	241	GACTATAAAAAGCTGTCAATTTCTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA	300
Qy	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Db	301	GACGGTGGCGTGGTTACTGTGTACCAACGAGATTCACAGTTTGACAGGATGGCTGTTTCATCTAC	360
Qy	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Db	361	AAAGTCAAGTTCATTGGCGTGAACTTTCTTCCGATGGACCTGTTATGCAAAAGAAGACA	420
Qy	141	MetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Db	421	ATGGGCTGGGAAGCCAGCACTCAGAGGTTTGTATCTCTCGTATGGCGTGTGTTGAAGAAGAGAG	480
Qy	161	IleHisLeuAlaLeuLysLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
Db	481	ATTCAATAGGCTCTGNAAGCTGAAAGACCGTGGTCAATTACCTAGTTGAATTCAAAAGTATT	540
Qy	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Db	541	TACATGGCAAAAGAAGCCTGTGCAGCTACCGAGGTACTATGTGTGATCTCCAAACTGGAT	600
Qy	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Db	601	ATAAACGAAGCCACACGAAGACTATACATCTGTTGAGCAGTATGAAGACCCGAGGAGCGC	660
Qy	221	HisHisLeuPheLeu	225
Db	661	CACCATCTGTCTCTT	675

RESULT 12

```

US-09-999-745-66
; Sequence 66, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 66
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-866-538-11

Alignment Scores:
Pred. No.: 6.12e-143 Length: 859
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-006-922A-12 (1-225) x US-09-866-538-11 (1-859)

Qy	1	MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Db	54	ATGAGGCTCTTCCAAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATCGAAGGA	113
Qy	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Db	114	ACGGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGGAGGGGCCATACGAAGGC	173
Qy	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Db	174	CACATACCGTAAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTTGGGTATATT	233
Qy	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Db	234	TTGTCAACCAAAATTTCAAGTATGAAGCAAGGATATGTCAAGCACCTTCCGACATACCA	293
Qy	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Db	294	GACTATAAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA	353
Qy	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Db	354	GACGGTGGCGTCGTACTGTAAACCCAGGATTCAGTTTTCAGGATGGCTGTTTCATCTAC	413
Qy	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Db	414	AAGGTCAAGTTCATTTGGCGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAGACA	473
Qy	141	IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle	180
Db	534	ATTCAATAGGCTCTGAAGCTGAAGCAGCGTGTGATCTCTCGTATGGCGTGTGAAAGGAGAG	533
Qy	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Db	594	TACATGCAAAAGAGCTGTGCAGCTGAAGCAGCGTGTGATCTCTCGTATGGCGTGTGAAAGGAGAG	533
Qy	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Db	654	ATAACAAGCCCAACGAAGACTATACAACTCGTTGAGCAGTATGAAGAACCGGAGGACGC	713
Qy	221	HisHisLeuPheLeu 225	
Db	714	CACCATCTGTTCCCT 728	

RESULT 14

US-09-794-308-11
; Sequence 11, Application US/09794308
; Publication No. US2003017091A1
; GENERAL INFORMATION:

; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSUEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-794-308-11

Alignment Scores:
Pred. No.: 6.12e-143 Length: 859
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-006-922A-12 (1-225) x US-09-794-308-11 (1-859)

Qy	1	MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Db	54	ATGAGGCTCTTCCAAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATCGAAGGA	113
Qy	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Db	114	ACGGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGGAGGGGCCATACGAAGGC	173
Qy	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Db	174	CACATACCGTAAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTTGGGTATATT	233
Qy	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Db	234	TTGTCAACCAAAATTTCAAGTATGAAGCAAGGATATGTCAAGCACCTTCCGACATACCA	293
Qy	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Db	294	GACTATAAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA	353
Qy	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Db	354	GACGGTGGCGTCGTACTGTAAACCCAGGATTCAGTTTTCAGGATGGCTGTTTCATCTAC	413
Qy	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Db	414	AAGGTCAAGTTCATTTGGCGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAGACA	473
Qy	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Db	474	ATGGCTGGGAAGCCAGCAGCTGAGCGTGTGATCTCTCGTATGGCGTGTGAAAGGAGAG	533
Qy	161	IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle	180
Db	534	ATTCAATAGGCTCTGAAGCTGAAGCAGCGTGTGATCTCTCGTATGGCGTGTGAAAGGAGAG	533
Qy	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Db	594	TACATGCAAAAGAGCTGTGCAGCTGAAGCAGCGTGTGATCTCTCGTATGGCGTGTGAAAGGAGAG	533
Qy	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Db	654	ATAACAAGCCCAACGAAGACTATACAACTCGTTGAGCAGTATGAAGAACCGGAGGACGC	713

Qy 221 HisHisLeuPheLeu 225
Db 714 CACCATCTGTTCTT 728

RESULT 15

US-09-865-291-11
; Sequence 11, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSIEN, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIO-METRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGEN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54) .. (731)
US-09-865-291-11

Alignment Scores:
Pred. No.: 6.12e-143 Length: 859
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-10-006-922A-12 (1-225) x US-09-865-291-11 (1-859)

Qy 1 MetArgSerLysAenValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 54 ATGAGGCTTCCAGAGATGTTATCAGGAGTTCATGAGGTTTAAGGTTCCGATGGAAGGA 113
Qy 21 ThrValAenGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 114 ACGGTCATAGGCGACGAGTTTGAATAAGCGGCGAAGGAGAGGGGCCATACGAAGGC 173
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 174 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTGCTTGGGATATT 233
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 234 TTGTCACACACAAATTCAGTATGGAAGCAGGATATATGTCAAGCACCCCTGCCGACATACCA 293
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 294 GACTATAAAAGCTGTCATTTCTCGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 353
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 354 GACGGTGGCGTCGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 413
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 414 AAGGTCAAGTTCATTTGGCGTGAACCTTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 473
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 474 ATGGGCTGGGAAGCAGCAGCTAGCGTTTGTATCCTCGTGATGGCGTGTGAAAGGAGAG 533
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Db 534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 593

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 594 TACATGGCAAGAAGGCTGTGCAGCTACCAAGGTACTACTATGTTGACTCCAAACTGGAT 653
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 654 ATAACAAGCCACAACCAAGAGACTATACAATCGTTTGAGCAGTATGAAAGAACCGAGGACGC 713
Qy 221 HisHisLeuPheLeu 225
Db 714 CACCATCTGTTCTT 728

Search completed: July 1, 2005, 21:04:50
Job time : 617 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 1, 2005, 18:23:08 ; Search time 186 Seconds
(without alignments)
1799.367 Million cell updates/sec

Title: US-10-006-922A-12
Perfect score: 1214
Sequence: 1 MRSSKNVKEFMRFKVRMEG.....EDYIVQEYRTEGRHHLFL 225

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO pool/US10006922/runat 30062005 110617 19116/app query.fasta_1.391
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10006922.cgn 1 1 69 @runat 30062005 110617 19116 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/1/ina/5A COMB.seq:
2: /cgn2_6/ptodata/1/ina/5B COMB.seq:
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length DB ID	Description
1	1214	100.0 678 3	US-09-459-956-6 Sequence 6, Appli
2	1210	99.7 723 4	US-10-152-296-1 Sequence 1, Appli
3	729.5	60.1 699 3	US-09-459-956-5 Sequence 5, Appli
4	667	54.9 801 3	US-09-459-956-7 Sequence 7, Appli
5	495	40.8 690 3	US-09-459-956-2 Sequence 2, Appli
6	486.5	40.1 1079 3	US-09-609-161B-15 Sequence 15, Appl
7	486.5	40.1 1079 4	US-09-626-581D-64 Sequence 64, Appl
8	486.5	40.1 1079 4	US-09-415-765B-64 Sequence 64, Appl
9	486.5	40.1 1079 4	US-09-626-580C-64 Sequence 15, Appl
10	486.5	40.1 1085 3	US-09-277-716-15 Sequence 4, Appli
11	485.5	40.0 696 3	US-09-459-956-4 Sequence 1, Appli
12	485.5	40.0 720 4	US-09-839-650-1 Sequence 1, Appli

13	477.5	39.3	696	3	US-09-459-956-3	Sequence 3, Appli
14	474.5	39.1	1021	4	US-09-839-650-2	Sequence 2, Appli
15	474.5	39.1	1104	3	US-09-277-716-30	Sequence 30, Appl
16	474.5	39.1	1104	3	US-09-609-161B-30	Sequence 30, Appl
17	474.5	39.1	1279	3	US-09-277-716-31	Sequence 31, Appl
18	474.5	39.1	1279	3	US-09-609-161B-31	Sequence 31, Appl
19	468.5	38.6	1482	4	US-09-977-897-1	Sequence 1, Appli
20	218.5	18.0	1559	3	US-09-049-475-6	Sequence 6, Appli
21	215.5	17.8	717	4	US-09-023-946B-3	Sequence 3, Appli
22	215.5	17.8	4196	4	US-09-453-313-1	Sequence 1, Appli
23	215.5	17.8	4199	3	US-09-204-117B-1	Sequence 1, Appli
24	215.5	17.8	6232	4	US-09-796-575-1	Sequence 19, Appl
25	215.5	17.8	7455	4	US-09-220-557-19	Sequence 23, Appl
26	215.5	17.8	7686	3	US-09-502-710-23	Sequence 26, Appl
27	215.5	17.8	7686	3	US-09-502-710-26	Sequence 23, Appl
28	215.5	17.8	7686	3	US-09-502-711-23	Sequence 26, Appl
29	215.5	17.8	7686	3	US-09-502-711-26	Sequence 26, Appl
30	215.5	17.8	7686	4	US-09-565-616A-2	Sequence 2, Appli
31	215.5	17.8	7687	3	US-09-502-710-24	Sequence 24, Appl
32	215.5	17.8	7687	3	US-09-502-711-24	Sequence 24, Appl
33	212.5	17.5	717	2	US-08-818-604-31	Sequence 31, Appl
34	212.5	17.5	717	2	US-09-346-946-31	Sequence 31, Appl
35	212.5	17.5	717	4	US-09-023-946B-21	Sequence 21, Appl
36	212.5	17.5	764	2	US-08-818-604-30	Sequence 30, Appl
37	212.5	17.5	764	3	US-08-819-612-21	Sequence 21, Appl
38	212.5	17.5	764	4	US-09-346-946-30	Sequence 30, Appl
39	212.5	17.5	764	4	US-09-872-364-21	Sequence 21, Appl
40	211.5	17.4	1665	2	US-08-771-850A-1	Sequence 1, Appli
41	211.5	17.3	716	1	US-08-337-915A-1	Sequence 1, Appli
42	210.5	17.3	716	1	US-08-753-143-1	Sequence 1, Appli
43	210.5	17.3	716	2	US-08-792-553-1	Sequence 1, Appli
44	210.5	17.3	716	3	US-08-753-144-1	Sequence 1, Appli
45	210.5	17.3	716	3	US-09-094-359-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-459-956-6
; Sequence 6, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, Iii, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR FILING DATE: 08/765,860
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp
US-09-459-956-6

Alignment Scores:
Pred. No.: 1.71e-160 Length: 678
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 3 Gaps: 0

US-10-006-922A-12 (1-225) x US-09-459-956-6 (1-678)

```
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGGTCTTCCAAGATGTTATCAAGAGTTTCATGAGTTTAAGGTTCCGATGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGGTCAATGGCGACGAGTTTGAATAAGAGCGAAGGAGGAGGCCATACGAAGGC 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 CACATACCGTAAAGCTTAAAGTTAACCAAGGGGGAGCTTTGCCATTTGCTTGGGATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATAAAAGCTGTCAATTTCTTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGCTGGCGTCTACTTACTTAACCCAGATTTCCAGTTTGCAAGATGGCTGTTTCATCTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 AAGGTCAAGTTTCATTGGCGTGAACCTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ATGGGCTGGGAAGCAGCAGCTAGCGTTTGTTATCCTCGTATGGCGTGTGAAGGAGAG 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATTCATAAAGGCTCTGAAGCTGAAGACGGTGGTCAATTACCTAGTTGAATTCAAAGATATT 540

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 541 TACATGCAAGAGAGCCTGTGCAGCTACCAAGGTACTACTATGTTGATCTCCAACCTGGAT 600

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 601 ATAACAAGCCACACGAAGACTATACAATCGTTGAGCAGTATGAAGAACCGAGGACGC 660

Qy 221 HisHisLeuPheLeu 225
Db 661 CACCATCTGTTCCTT 675
```

RESULT 2

```
US-10-152-296-1
; Sequence 1, Application US/10152296
; Patent No. 6723537
; GENERAL INFORMATION:
; APPLICANT: Peelle, Beau
; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
; FILE REFERENCE: 021044-000110US
; CURRENT APPLICATION NUMBER: US/10/152,296
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/291,871
; PRIOR FILING DATE: 2001-05-18
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 723
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mammalian
; OTHER INFORMATION: codon-optimized variant (dsRED) of Discosoma sp.
; OTHER INFORMATION: "red" red fluorescent protein (RFP)
; FEATURE:
; NAME/KEY: CDS
```

```
; LOCATION: (1)..(723)
; OTHER INFORMATION: dsRED
US-10-152-296-1
```

```
Alignment Scores: 6.93e-160 Length: 723
Pred. No.: 1210.00 Matches: 224
Score: 100.00% Conservative: 1
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 99.56% Indels: 0
Query Match: 99.67% Gaps: 0
DB: 4
```

US-10-006-922a-12 (1-225) x US-10-152-296-1 (1-723)

```
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 4 GTCCGCTCTCTCCAAGAACGTCATCAAGGAGTTTCATCGCTTCAAGGTGGCATGGAGGC 63

Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 64 ACCGTGACCGCCACGAGTTCGAGATCGAGGCGGAGGCGGCGCCCTACGAGGC 123

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 124 CACAACACCGTGAAGCTGAAGGTGACCAAGGGCGGCGCCCTTCGCCCTTCGCTCGGACATC 183

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 184 CTGTCCTCCCGAGTTCCAGTACGCGCTCCAAGGTGTACGTGAAGCACCCTCCGCGACATCCCC 243

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 244 GACTACAGAAGAGTGTCTTCCCGAGGCGCTCAAGTGGGAGCGCGTGATGAACCTTCGAG 303

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 304 GACGGCGGCTGGTGACCGTGCAGCCAGGACTCTCTCCCTCGCAGGACGGCTGCTTCATCTAC 363

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 364 AAGTGAAGATTTCATCGCGGTGAACCTTCCCTCCGAGCGGCCCGGTAATGAGAAGAAGACC 423

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 424 ATGGGCTGGGAGGCGCTCCACCGAGCGCTGTACCCCGCGCGCGCTGCTGAAGGCGGAG 483

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 484 ATCCACAAGCGCTGAAGCTGAAGGACGCGCGCCACTACTCTGGTGAGTTCAAGAGTATC 543

Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Db 544 TACATGGCAAGAGCGCCGCTGCGAGCTGCCCGGCTACTACTACGTGACTCCAAGCTGGAC 603

Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Db 604 ATCACTCCCAACACGAGGACTACACCATCGTGGAGCAGTACGAGCGCACCGAGGCGCGC 663

Qy 221 HisHisLeuPheLeu 225
Db 664 CACCACCTGTTCCTG 678
```

RESULT 3

```
US-09-459-956-5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Teien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
```

```

; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

Alignment Scores:
Pred. No.: 1,95e-92 Length: 699
Score: 729.50 Matches: 133
Percent Similarity: 76.00% Conservative: 38
Best Local Similarity: 59.11% Mismatches: 53
Query Match: 60.09% Indels: 1
DB: 3 Gaps: 1

US-10-006-922A-12 (1-225) x US-09-459-956-5 (1-699)

Qy 1 MetArgSerSerLysAsnValLleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGAGTTGTTCCAGAGAGTGTGATCAAGGAAGAAATGTTGATCGATCTTCATCTCGGAAGGA 60

Qy 21 ThrValAsnGlyHisGluPheGluLeuGluGlyGluGlyArgProTyrGluGly 40
Db 61 ACGTTCATAGGCGACTACTTTGAATTAAGGAAGAAAGGACAGCCCTAATGAAGGC 120

Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 ACCAATACCGTCACGCTCGAGGTTACCAAGGTTGAGCCCTCTGCCATTTGGTTGCATATT 180

Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTGCCACCAATTTCAAGTATGGAACAAGGCAATTTGTCCACCCCTGACCAACATACAT 240

Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GATTATCTAAAGCTGTCTATTTCCGAGGATATATACATGGGACCGTCCATGCATCTTTGAA 300

Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Db 301 GACGGTGGCTTGTGTGTATCACCAATGATATCAGTTTGACAGGCAACTGTTCTACTAC 360

Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Db 361 GACATCAAGTTCACCTGGCTTGAACCTTCTCCAAATGGACCCGTTGTGCAGAGAAGACA 420

Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Db 421 ACTGGCTGGGAACCGGACGACCTGAGCGTTGTATCTCTGTGATGGTGTGTTGATAGGAGC 480

Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Db 481 ATCCATCATGCTCTGACAGTTGAAGGAGGTGTTGATACCGATCGATGACATTAACACTGTT 540

Qy 181 TyrMetAlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 199
Db 541 TACAGGGCCAAAGAGCGCGCTTGAAGATGCGAGGATATCACTATGTTGACACCAAACTG 600

Qy 200 AspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGly 219
Db 601 GTTATATGGAACACGACCAAGAAATTCATGAAGTTGAGGAGCATGAATAATCGCGTTGCA 660

Qy 220 ArgHisHisLeuPhe 224
Db 661 CGCCACCATCCGTTTC 675

US-09-459-956-7

Alignment Scores:
Pred. No.: 1,45e-83 Length: 801
Score: 667.00 Matches: 121
Percent Similarity: 73.76% Conservative: 42
Best Local Similarity: 54.75% Mismatches: 58
Query Match: 54.94% Indels: 0
DB: 3 Gaps: 0

US-10-006-922A-12 (1-225) x US-09-459-956-7 (1-801)

Qy 3 SerSerLysAsnValLleLysGluPheMetArgPheLysValArgMetGluGlyThrVal 22
Db 121 ACCACCATGGTGTGATTAACCCAGACATGAAGATTAAAGCTGAAGATGAAGGAATGTA 180

Qy 23 AsnGlyHisGluPheGluLeuGluGlyGluGlyArgProTyrGluGlyHisAsn 42
Db 181 AACGGGATGCTTTGTGATCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240

Qy 43 ThrValLysLysValThrLysGlyGlyProLeuProPheAlaTrpAspIleLeuSer 62
Db 241 ACTTTAAACCTCGAAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300

Qy 63 ProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIleProAspTyr 82
Db 301 AACGCGTTCAGTACGGAACACAGAGCATTCGACAAATACCCAGACGATATAGCAGACTAT 360

Qy 83 LysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGluAspGly 102
Db 361 TTCAGCAGTCTGTTCCGAGGAGATATCTCGGAAAGAGACCATGACTTTTGAAGACAAA 420

Qy 103 GlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyrLysVal 122
Db 421 GGCATCTCAAGTGAAGAGTGAACATGAAGCATGGAGGAGAGAGAGAGAGAGAGAGAGAG 480

Qy 123 LysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThrMetGly 142
Db 481 CGTTTTGATGGATGAACCTTTCTCCCAATGGTCCGCTTATGCAAGAAAAAATTTTGAAG 540

Qy 143 TrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGluIleHis 162
Db 541 TGGACCATCTCCACTGAGATTATGATGCTGGTGTGATGAGTCTGCTGGAGATATTAGC 600

Qy 163 LysAlaLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 182
Db 601 CATTTCTGTTGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660

Qy 183 AlaLysLysProValGlnLeuLeuProGlyTyrTyrTyrValAspSerLysLeuAspIleThr 202
```


Qy 57 AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValThrValLysHisPro 76
Db 436 GCATTGTGATATTGTCACCAAGCTTTTCAATATATGCGCAACCGTACTTTTCCAGAAATATCCG 495
Qy 77 AlaAspIleProAspTyrLysLysLeuSerPhePheProGluGlyPheLysTyrGluArgVal 96
Db 496 AATGATATATACAGATTATTTATACATCATTTCCAGCAGGATTATGATGAACGAACA 555
Qy 97 MetAsnPheGluAspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGly 116
Db 556 TTACGTTACGAAGATGCGGACTTTGTAATTCGTTACAGATATAAATTTTAAATAGAAGAC 615
Qy 117 CysPheIleTyrLysValLysPheIleGlyValAsnPhePheProSerAspGlyProValMet 136
Db 616 AAGTTCGCTACAGAGTGAATACAAAGGTAGTAACTTCCAGATGATGGTCCCGTCATG 675
Qy 137 GlnLysLysThrMetGlyTyrGluAlaSerThrGluArgLeuTyrProAspGlyVal 156
Db 676 CAGAAAGACTATCTAGGAATAGACCTTTCATTGAGCCATGATACATGAATATGGCGTC 735
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGlu 176
Db 736 TTGTCGCGAAGTAATCTTGTCTATAAACTAACTCTGGAAATATTTATTCATGTCAC 795
Qy 177 PheLysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrVal 195
Db 796 ATGAAACATTAAATGAAGTCGAAAGGTAGTAAAGGAGTTTCCTCGTATCATTTATT 855
Qy 196 AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu 215
Db 856 CAACATCGTTGGAAAGACT---TAGTGAAAGACGGGGGTTTCGTTGAACACATGAG 912

RESULT 7

US-09-626-581D-64
; Sequence 64, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 64
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-626-581D-64

Alignment Scores:
Pred. No.: 5.18e-58 Length: 1079
Score: 486.50 Matches: 97
Percent Similarity: 64.09% Conservative: 44
Best Local Similarity: 44.09% Mismatches: 72
Query Match: 40.07% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-12 (1-225) x US-09-626-581D-64 (1-1079)

Qy 2 ArgSerLysAsnValIleLys-----GluPheMetArgPheLysVal 16
Db 256 AAGATGATGATAAACAATATTGGAAGACACACTGTTTACAGAAGTAATGTCGTATAAAGTA 315

Qy 17 ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArg 36
Db 316 AATCTGGAAGGAATGTGAAACAACCATGTTTTCACATGGAGGGTTCCGCGCAAGGAAT 375
Qy 37 ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPhe 56
Db 376 ATTTTATTCGCAATCACTACCTGGTTTCAGATTCGTTGTCACGAAGGGGCCCTCCCTTTT 435
Qy 57 AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValThrValLysHisPro 76
Db 436 GCATTGTGATATTGTCACCAAGCTTTTCAATATATGCGCAACCGTACTTTTCCAGAAATATCCG 495
Qy 77 AlaAspIleProAspTyrLysLysLeuSerPhePheProGluGlyPheLysTyrGluArgVal 96
Db 496 AATGATATATACAGATTATTTATACATCATTTCCAGCAGGATTATGATGAACGAACA 555
Qy 97 MetAsnPheGluAspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGly 116
Db 556 TTACGTTACGAAGATGCGGACTTTGTAATTCGTTACAGATATAAATTTTAAATAGAAGAC 615
Qy 117 CysPheIleTyrLysValLysPheIleGlyValAsnPhePheProSerAspGlyProValMet 136
Db 616 AAGTTCGCTACAGAGTGAATACAAAGGTAGTAACTTCCAGATGATGGTCCCGTCATG 675
Qy 137 GlnLysLysThrMetGlyTyrGluAlaSerThrGluArgLeuTyrProAspGlyVal 156
Db 676 CAGAAAGACTATCTAGGAATAGACCTTTCATTGAGCCATGATACATGAATATGGCGTC 735
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLysAspGlyGlyHisTyrLeuValGlu 176
Db 736 TTGTCGCGAAGTAATCTTGTCTATAAACTAACTCTGGAAATATTTATTCATGTCAC 795
Qy 177 PheLysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrVal 195
Db 796 ATGAAACATTAAATGAAGTCGAAAGGTAGTAAAGGAGTTTCCTCGTATCATTTATT 855
Qy 196 AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu 215
Db 856 CAACATCGTTGGAAAGACT---TAGTGAAAGACGGGGGTTTCGTTGAACACATGAG 912

RESULT 8

US-09-415-765B-64
; Sequence 64, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 64
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-415-765B-64

Alignment Scores:
Pred. No.: 5.18e-58 Length: 1079
Score: 486.50 Matches: 97
Percent Similarity: 64.09% Conservative: 44
Best Local Similarity: 44.09% Mismatches: 72
Query Match: 40.07% Indels: 7
DB: 4 Gaps: 3

```
US-10-006-922A-12 (1-225) x US-09-415-765B-64 (1-1079)
QY 2 ArgSerSerLysAsnValIleLys-----GluPheMetArgPheLysVal 16
Db 256 AAGATGATGTAACAATAATGGAAGAACACTTGTGTTTACAAGAAGTAATGTCGTATAAAGTA 315
QY 17 ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArg 36
Db 316 AATCTGGAAGGAATTGTAACAACCACTGTTTACAATGGAGGGTTGCGGCAAGGAAT 375
QY 37 ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyProLeuProPhe 56
Db 376 ATTTTATTCGGCAATCAACTGTTTCAGATTCGTGTCAGAAAGGGCCCACTGCCCTTT 435
QY 57 AlaTrpAspIleLeuSerProGlnPheGlnThrValThrGlnAspSerSerLysHisPro 76
Db 436 GCATTTGATATGTTGTCACCAAGCTTTTCAATATGCAACCGTACTTTCACGAAATATCCG 495
QY 117 CysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMet 136
Db 616 AAGTTCGTCTACAGATGGAATACAAAGGTAGTAACCTCCAGATGATGGTCCCGTCATG 675
QY 137 GlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyVal 156
Db 676 CAGAAGACTATCTTAGAATAGACCTTCAATTCAGCCATGACATGAATAATGCGGTC 735
QY 196 AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu 215
Db 856 CAACATCGTTTGGAAAGACT--TACGTAGAAGACGGGGGTTTCGTTGAACACATGAG 912

RESULT 9
US-09-626-580C-64
; Sequence 64, Application US/09626580C
; Patent No. 6562617
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; APPLICANT: Bogenberger, Jakob M.
; APPLICANT: Peele, Beau R.
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 64
; LENGTH: 1079
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
```

US-09-626-580C-64

Alignment Scores: 5.18e-58 Length: 1079
Pred. No.: 486.50 Matches: 97
Score: 64.09% Conservative: 44
Best Similarity: 44.09% Mismatches: 72
Query Match: 40.07% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-12 (1-225) x US-09-626-580C-64 (1-1079)

```
QY 2 ArgSerSerLysAsnValIleLys-----GluPheMetArgPheLysVal 16
Db 256 AAGATGATGTAACAATAATGGAAGAACACTTGTGTTTACAAGAAGTAATGTCGTATAAAGTA 315
QY 17 ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArg 36
Db 316 AATCTGGAAGGAATTGTAACAACCACTGTTTACAATGGAGGGTTGCGGCAAGGAAT 375
QY 37 ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyProLeuProPhe 56
Db 376 ATTTTATTCGGCAATCAACTGTTTCAGATTCGTGTCAGAAAGGGCCCACTGCCCTTT 435
QY 57 AlaTrpAspIleLeuSerProGlnPheGlnThrValThrGlnAspSerSerLysHisPro 76
Db 436 GCATTTGATATGTTGTCACCAAGCTTTTCAATATGCAACCGTACTTTCACGAAATATCCG 495
QY 77 AlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgVal 96
Db 496 AATGATATATCAGATTAATTTTACAATCATTTCCAGCAGGATTTATGTATGAACGAACA 555
QY 97 MetAsnPheGluAspGlyGlyValValThrValThrGlnAspSerSerLysLeuGlnAspGly 116
Db 556 TTACGTTACGAAGATGCGGACTTGTGTTTCAATAAATTCGTTCAAGATATAATTTAATAGAAGAC 615
QY 117 CysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMet 136
Db 616 AAGTTCGTCTACAGATGGAATACAAAGGTAGTAACCTCCAGATGATGGTCCCGTCATG 675
QY 137 GlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyVal 156
Db 676 CAGAAGACTATCTTAGAATAGACCTTCAATTCAGCCATGACATGAATAATGCGGTC 735
QY 157 LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGlu 176
Db 736 TTGTCGGCGGAAGTAATTCCTGCTATAAATCTCTGGAAATATATTTATCATGTAC 795
QY 177 PheLysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrVal 195
Db 796 ATGAAACATTAATGAAGTCGAAAGGTAGTAGTAAGGAGTTTCCTTCGTATCATTTTATT 855
QY 196 AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu 215
Db 856 CAACATCGTTTGGAAAGACT--TACGTAGAAGACGGGGGTTTCGTTGAACACATGAG 912

RESULT 10
US-09-277-716-15
; Sequence 15, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-10-01
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-06-15
; EARLIER FILING DATE: 1998-03-27
```

; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-277-716-15

Alignment Scores:
Pred. No.: 5,23e-58 Length: 1085
Score: 486.50 Matches: 97
Percent Similarity: 64.09% Conservative: 44
Best Local Similarity: 44.09% Mismatches: 72
Query Match: 40.07% Indels: 7
DB: 3 Gaps: 3

US-10-006-922A-12 (1-225) x US-09-277-716-15 (1-1085)

```
Qy 2 ArgSerSerLysAsnValIleLys-----GluPheMetArgPheLysVal 16
Db 256 AAGTGTAGTAAACAAATATTGAAGACACTGTTTACAAAGAAATGTCGTATAAGTA 315
Qy 17 ArgMetGluGlyThrValAsnGlyHisGluPheGluGlyGluGlyArg 36
Db 316 AATCTGGAAGGAATTGAAACACCATGTTTTCATCAATGAGGGTTCGGCAAGGAAT 375
Qy 37 ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPhe 56
Db 376 ATTTTATTCGCAATCAACTGGTTTCAGATTGCTGTCAGAAAGGGGCCCACTGCCCTTT 435
Qy 57 AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisPro 76
Db 436 GCATTGTATATGTGTACACAGCTTTTCAATATGCAACCGCTACTTTCCAGAAATATCCG 495
Qy 77 AlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgVal 96
Db 496 AATGATATATCAGATTATTTATACAAATCATTTCCAGCAGGATTATGATGACGAACA 555
Qy 97 MetAsnPheGluAspGlyValValThrValThrGlnAspSerSerLeuGlnAspGly 116
Db 556 TTACGTTTACGAAGATGGCGACTTGTTCAAATTCGTTCCAGATATAAAATTTAATAGAAGAC 615
Qy 117 CysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMet 136
Db 616 AAGTTCGTCTACAGAGTGGGAATACAAAGGTAGTAACCTCCAGATGATGGTCCCGTCATG 675
Qy 137 GlnLysLysThrMetGlyTyrGluAlaSerThrGluArgLeuTyrProArgAspGlyVal 156
Db 676 CAGAGACTACTCTTAGGATAGACCTTCTTTGAGCCCATGATACATGAATAGGCGTC 735
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGlu 176
Db 736 TTGGTCGCGGAAGTAATTCCTGTCTATAAACTAAACTCTGGAAATATTATTCATGTCAC 795
Qy 177 PheLysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrVal 195
Db 796 ATGAAACATTAATGAAGTCAAAGGTGATGAAGGAGTTTCTTCGTATCATTTATT 855
Qy 196 AspSerLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu 215
Db 856 CAACATCGTTTGGAAAGACT--TAGTAGAAGACGGGGGTTTCGTTGAACATGAG 912
```

RESULT 11

US-09-459-956-4
; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.

; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: OPTICAL METHODS
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIORITY FILING DATE: 1999-05-08
; PRIORITY FILING DATE: 1999-05-08
; PRIORITY FILING DATE: 1999-05-08
; PRIORITY FILING DATE: 1999-06-07
; PRIORITY FILING DATE: 1999-06-07
; PRIORITY FILING DATE: 1999-06-07
; PRIORITY FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
; ORGANISM: Zoanthus sp
US-09-459-956-4

Alignment Scores:
Pred. No.: 3,54e-58 Length: 696
Score: 485.50 Matches: 95
Percent Similarity: 64.71% Conservative: 37
Best Local Similarity: 46.57% Mismatches: 67
Query Match: 39.99% Indels: 5
DB: 3 Gaps: 3

US-10-006-922A-12 (1-225) x US-09-459-956-4 (1-696)

```
Qy 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Db 1 ATGGCTCATTTCAAAGCACGGCTTAAAGAAAGAAATGCAATGAAATACCATCGAAGGG 60
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db 61 TGGCTCAACGACATATAATTTGTGATCACGGCGGAGGCGATTGGATATCCGTTCAAAGGG 120
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Db 121 AAACAGACTTAAATCTGTGTGTGATCGAAGGGGGACCATTCCTTTCGAAAGACATA 180
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Db 181 TTGTGACGTGGCTTTAAGTACGAGACAGGATTTTCACTGAATATCTCTCAAGACATAGTA 240
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Db 241 GACTATTTCAAGAACTCGTCTCTGCTGGATATACATGGGGCAGGCTTTTCTCTTTTGGAG 300
Qy 101 AspGlyGlyValValThrValThrGlnAsp-----SerSerLeuGlnAspGlyCysPhe 118
Db 301 GATGGACAGCTCTGCATATGCAATGATAGATAAACAAGTGTGTCGAAAGAAACTGCATT 360
Qy 119 IleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLys 138
Db 361 TATCATTAAGACATATTAAATGAATGAATTTTCTCTGCTGATGGACCTGTGATGAAAGAG 420
Qy 139 LysThrMetGlyTyrGluAlaSerThrGluArgLeu-----TyrProArgAspGlyVal 156
Db 421 ATGCAACTAACTGGGAGAGCATCTCGCAGAGAGATCATGCCAGTACCTTAAGCAGGGGATA 480
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGlu 176
Db 481 CTGAAAGGGGATGCTCCATGTACCTCTCTTCTGAAGGATGGTGGCGCTTACCGGTGCCAG 540
Qy 177 PheLysSerIleTyrMetAlaLysLysLys---ProValGlnLeuProGlyTyrTyrVal 195
Db 541 TTCGACACAGTTTACAAAGCAAGAGTCTGTGCCAAGTAAGATGCCGGAGTGCGCACTTCATC 600
Qy 196 AspSerLysLeu 199
Db 601 CAGCATTAAGCTC 612
```

```

RESULT 12
US-09-839-650-1
; Sequence 1, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; FILE OF INVENTION: Fluorescent Protein
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Humanized R. mulleri polynucleotide
; NAME/KEY: misc feature
; LOCATION: (1)..(720)
; OTHER INFORMATION: Humanized DNA sequence
US-09-839-650-1

Alignment Scores:
Pred. No.:          3,74e-58      Length:          720
Score:              485.50        Matches:         97
Percent Similarity: 64.22%       Conservative:    43
Best Local Similarity: 44.50%     Mismatches:     71
Query Match:       39.99%        Indels:         7
DB:                4            Gaps:           3

US-10-006-922A-12 (1-225) x US-09-839-650-1 (1-720)

Qy   4 SerLysAsnValIleLys-----GluePheMetArgPheLysValargMet 18
Dbb  7 AGCAAGCAGATCTCTGAAGAACACCTGCTGCAGGAGGTGATGAGCTACAAGGTGAACCTG 66
Qy   19 GluGlyThrValAsnGLYHisGluPheGluLeuLIEgluGlyGluGlyGluGlyArgProTyr 38
Dbb  67 GAGGGCATCGTGAAACCACCGTGTTCACCATGGAGGGCTCGGGCAAGGGCAACATCCGTG 126
Qy   39 GluGlyHisAsnThrValLysLeuLysValThrlYsGlyGlyProLeuProPheAlaTrp 58
Dbb  127 TTCGGCAACCAAGCTGGTGCAGATCGCTGACCAAGGGCGCCCCCTGCCCTTGGCCTTC 186
Qy   59 AspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaasp 78
Dbb  187 GACATCGTAGCCCCCGCTTCCAGTACGGCAACCGCACCTTCCACCAGATACCCCAACGAC 246
Qy   79 IleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsn 98
Dbb  247 ATCAGCGACTACTTCATCCAGAGCTTCCC CGCGGCTTCATGTACGAGCGCACCTGGCG 306
Qy   99 PheGluaspGlyGlyValvalThrValThrGlnaspSerSerLeuGlnaspGlyCysPhe 118
Dbb  307 TACGAGGACGGCGGCTGTGTGAGATCGCGACGACATCAACCTGATCGAGGACAAGTTTC 366
Qy   119 IleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLys 138
Dbb  367 GTGTACCGGTGGAGTACAAGGCGACAACTTCCC CGACGACGGCCCCCTGATGCGAAG 426
Qy   139 LysThrMetGlyTrpGluAlaserThrGluArgLeuTyrProArgaspGlyValLeuLys 158
Dbb  427 ACCATCTCTGGGCATCGAGCCCGAGCTTCGAGGCCATGTATCATCAAACGGCGTGTGGTG 486
Qy   159 GlyGluIleHisLysAlaLeuLysLeuLysaspGlyGlyHisTyrLeuValGluPheLys 178
Dbb  487 GCGGAGGTGATCTCTGGTGTAACAAGCTGAACAGCGGCAAGTACTACTACGCTGCCACATGAAG 546
Qy   179 SerileTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrValAspSer 197

```


Db 421 ATGACAGATACTGGAGCCATCTCGAGAGATCATACCAGTACCTTAAGCGGGATA 480
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGlu 176
Db 481 TTGAAGGGGATGCTCCATGATCTCTCTGAGGATGTTGGCGGTTTACGGTGCCAA 540
Qy 177 PhelysSerIleTyrMetAlaLysLys---ProValGlnLeuProGlyTyrTyrVal 195
Db 541 TTCACACAGTTTACAAGCAAGTCTGTGCAAGAAAGATGCCGGAGTGCACCTTCATC 600
Qy 196 AspSerLysLeu 199
Db 601 CAGCATTAAGCTC 612

RESULT 14

US-09-839-650-2
; Sequence 2, Application US/09839650
; Patent No. 6645761
; GENERAL INFORMATION:
; APPLICANT: Stratagene
; TITLE OF INVENTION: Humanized Polynucleotide Sequence Encoding Renilla Mulleri Green
; Patent No. 6645761
; FILE REFERENCE: 25436/1755
; CURRENT APPLICATION NUMBER: US/09/839,650
; CURRENT FILING DATE: 2001-04-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1021
; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: exon
; LOCATION: (259)..(976)
US-09-839-650-2

Alignment Scores:
Pred. No.: 2,3e-56 Length: 1021
Score: 474.50 Matches: 96
Percent Similarity: 63.18% Conservative: 43
Best Local Similarity: 43.64% Mismatches: 74
Query Match: 39.09% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-12 (1-225) x US-09-839-650-2 (1-1021)

Qy 2 ArgSerLysAsnValIleLys-----GluPheMetArgPheLysVal 16
Db 256 AAGATGAGTAACAAATATTGAGAACACTGTTTACAGAAAGTAATGCTGATAAAGTA 315
Qy 17 ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArg 36
Db 316 AATCTGGAAGGAATTGAAACAACCATGTTTTCATATGAGGGTTCGCCCAACGGAAT 375
Qy 37 ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPhe 56
Db 376 ATTTTATTCGCAATCACTGTTTTCATATTCGTCACGAAAGGGGCCCACTGCCTTTT 435
Qy 57 AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisPro 76
Db 436 GCATTGATATGTCACCAAGCTTTTCAATATGCAACCGCTACTTTTCAGAAATATCCG 495
Qy 77 AlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTyrPgluArgVal 96
Db 496 AATGATATATCAGATTATTTATACATCATTTCCAGCAGGATTATGATGACGACAA 555
Qy 97 MetAsnPheGluAspGlyValValThrValThrGlnAspSerLysLeuGlnAspGly 116
Db 556 TTACGTTACGAAGATGGCGGCTGTTGMAATTCGTTTCAGATATAAATTTAATAGAGAC 615
Qy 117 CysPheIleTyrLysValLysPheIleGlyValAsnProSerAspGlyProValMet 136

Db 616 AAGTTCGTCTACAGAGTGGAAATACAAAGTAGTAACCTTCCAGATGATGGTCCGTCATG 675
Qy 137 GlnLysLysThrMetGlyTyrPgluAlaSerThrGluArgLeuTyrProArgAspGlyVal 156
Db 676 CAGAAGACTATCTTAGGAATAGAGCTTTCATTTGAAGCCATGTACATGAATAATGGGCTC 735
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGlu 176
Db 736 TTGTCGCGCAAGTAATCTTGTCTATAAACTAACTCTCGGAAATATTATTTCATGTCAC 795
Qy 177 PhelysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrVal 195
Db 796 ATGAACAATTAAATGAAGTCGAAAGGTGTAGTAAGGAGTTCCTTCGTATCATTTTATT 855
Qy 196 AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGlu 215
Db 856 CAACATCGTTTGGAAAAGACT---TACGTAGAAGACGGGGGTTTCGTTGAACAGCATGAG 912

RESULT 15

US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent in Ver. 2.0

SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)

US-09-277-716-30

Alignment Scores:
Pred. No.: 2,6e-56 Length: 1104
Score: 474.50 Matches: 97
Percent Similarity: 63.64% Conservative: 43
Best Local Similarity: 44.09% Mismatches: 73
Query Match: 39.09% Indels: 7
DB: 3 Gaps: 3

US-10-006-922A-12 (1-225) x US-09-277-716-30 (1-1104)

Qy 2 ArgSerLysAsnVal-----IleLysGluPheMetArgPheLysVal 16
Db 31 AAATGAAACCGCAACGATTTAAAGAACACTGGACTGAAGAGATTATGTGGCAAAAGCT 90
Qy 17 ArgMetGluGlyThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArg 36
Db 91 AGCGTTGAAGGAATCGTGAACAATCACGTTTTTTCATGGAAGGATTTTGAAGAGCAAT 150
Qy 37 ProTyrGluGlyHisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPhe 56
Db 151 GTATTATTTCGAAACCAATTCATGATCGGGTTCACAAAGGGAGGTCCGTTGCCATTTC 210
Qy 57 AlaTrpAspIleLeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisPro 76
Db 211 GCTTTCGATATTGTTTCCATAGCTTTTCCATACGGGANTCGCACTTTTCAGAAATACCCA 270

```
Qy 77 AlaAspIleProAspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgVal 96
Db 271 GACGACATTGGGACTACTTTGTTCAATCATTCCTCGGCTGGATTTTCTACGAAAGAAAT 330
Qy 97 MetAsnPheGluAspGlyGlyValValThrGlnAspSerSerLeuGlnAspGly 116
Db 331 CTACGCTTTGAAGATGGCCATTGTTGACATTGCTTCAGATATAAGTTTAGAAGATGAT 390
Qy 117 CysPheIleTyrLysValLysPheIleGlyValAsnPheProSerAspGlyProValMet 136
Db 391 AAGTTCACCTACAAAGTGGAGTATAGAGGCAACGGTTTCCTAGTAACGACCCGCGTATG 450
Qy 137 GlnLysLysThrMetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyVal 156
Db 451 CAAAAGCCCATCTCGGCATGGAGCCATCGTTGAGGTGGTCTACATGAACAGCGGCGTT 510
Qy 157 LeuLysGlyGluIleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGlu 176
Db 511 CTGGTGGCGAAGTAGATCTCGTTTACAACTCGAGTCAGGGAATATTACTCGTGCCAC 570
Qy 177 PheLysSerIleTyrMetAlaLysLysProVal---GlnLeuProGlyTyrTyrVal 195
Db 571 ATGAAAACGTTTACAGATCCAAAGGTGGAGTGAAGAATTCGCCGAATATCACTTTATC 630
Qy 196 AspSerLysLeuAspIleThrSerHisAsnGluAspTyrThrIleValGlnTyrGlu 215
Db 631 CATCATCGTCTGGAGAAACC--TACGTGGAGAGGAAGCTTCGTGGAAACACACAG 687
```

Search completed: July 1, 2005, 20:54:30
Job time : 189 secs

Db 241 GACTATATAAAGCTGTCTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Qy 301 GAGGTCGGCGTCTTACTGTAAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
Db 301 GAGGTCGGCGTCTTACTGTAAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGGTCGAAGTTCAATTTGGCGTGAACTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCGAAGTTCAATTTGGCGTGAACTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 421 ATGGGTCGGGAAGCCAGCACTAGCGTTTGTATCCTCGTGAAGCGGTTGTAAGGAGAG 480
Db 421 ATGGGTCGGGAAGCCAGCACTAGCGTTTGTATCCTCGTGAAGCGGTTGTAAGGAGAG 480
Qy 481 ATTCAATAGGCTCTGAAGCTGAAAGCGGTGGTCAATTAACCTAGTTGAATTCGAAGGATTT 540
Db 481 ATTCAATAGGCTCTGAAGCTGAAAGCGGTGGTCAATTAACCTAGTTGAATTCGAAGGATTT 540
Qy 541 TACATGGCAAGAGCCCTGTGCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAGAGCCCTGTGCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATAAAGCCACACGAAAGCTATACAAATCGTTGAGCAGTATGAAAGACCGAGGAGCG 660
Db 601 ATAAAGCCACACGAAAGCTATACAAATCGTTGAGCAGTATGAAAGACCGAGGAGCG 660
Qy 661 CACCATCTGTTCTTTAA 678
Db 661 CACCATCTGTTCTTTAA 678

RESULT 2
AX172854
LOCUS AX172854 678 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 6 from Patent WO0142211.
ACCESSION AX172854
VERSION AX172854.1 GI:14597903

SOURCE
ORGANISM
Discosoma sp.
Discosoma sp.
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.

REFERENCE
AUTHORS Tsien, R.Y. and Gonzalez, J.E.
TITLE Detection of transmembrane potentials by optical methods
JOURNAL Patent: WO 0142211-A 6 14-JUN-2001;
The Regents of the University of California (US)

FEATURES
source
1..678
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"

ORIGIN
Query Match 100.0%; Score 678; DB 6; Length 678;
Best Local Similarity 100.0%; Pred. No. 1.2e-189;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGTCCTTCAAGAATGTTATCAAGGATTTCAAGGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 ATGAGGTCCTTCAAGAATGTTATCAAGGATTTCAAGGTTTAAAGTTTCGCATGGAAGGA 60
Qy 61 ACAGTCATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGGAGGCCATACGAAGGC 120
Db 61 ACAGTCATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGGAGGCCATACGAAGGC 120
Qy 121 CACAATACCGTAAAGCTTAAAGGTAACCAAGGGGAGCCCTTGGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAAGGTAACCAAGGGGAGCCCTTGGCTTGGGATATT 180
Qy 181 TTGTACACCAATTTTCAGTATGGAAGCAGAGTATATGTCAAGCAACCCCTGCCGACATACCA 240
Db 181 TTGTACACCAATTTTCAGTATGGAAGCAGAGTATATGTCAAGCAACCCCTGCCGACATACCA 240

Qy 241 GACTATATAAAGCTGTCTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATATAAAGCTGTCTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Qy 301 GACGTCGGCGTCTTACTGTAAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGTCGGCGTCTTACTGTAAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGGTCGAAGTTCAATTTGGCGTGAACTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCGAAGTTCAATTTGGCGTGAACTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 421 ATGGGTCGGGAAGCCAGCACTAGCGTTTGTATCCTCGTGAAGCGGTTGTAAGGAGAG 480
Db 421 ATGGGTCGGGAAGCCAGCACTAGCGTTTGTATCCTCGTGAAGCGGTTGTAAGGAGAG 480
Qy 481 ATTCAATAGGCTCTGAAGCTGAAAGCGGTGGTCAATTAACCTAGTTGAATTCGAAGGATTT 540
Db 481 ATTCAATAGGCTCTGAAGCTGAAAGCGGTGGTCAATTAACCTAGTTGAATTCGAAGGATTT 540
Qy 541 TACATGGCAAGAGCCCTGTGCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAGAGCCCTGTGCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATAAAGCCACACGAAAGCTATACAAATCGTTGAGCAGTATGAAAGACCGAGGAGCG 660
Db 601 ATAAAGCCACACGAAAGCTATACAAATCGTTGAGCAGTATGAAAGACCGAGGAGCG 660
Qy 661 CACCATCTGTTCTTTAA 678
Db 661 CACCATCTGTTCTTTAA 678

RESULT 3
AX207715
LOCUS AX207715 678 bp DNA linear PAT 31-AUG-2001
DEFINITION Sequence 13 from Patent WO0157242.
ACCESSION AX207715
VERSION AX207715.1 GI:15422399

SOURCE
ORGANISM
Discosoma sp.
Discosoma sp.
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.

REFERENCE
AUTHORS Stack, J.H., Whitney, M., Cubitt, A.B. and Pollok, B.A.
TITLE Methods of protein destabilization and uses thereof
JOURNAL Patent: WO 0157242-A 13 09-AUG-2001;
Aurora Biosciences Corporation (US)

FEATURES
source
1..678
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"

ORIGIN
Query Match 100.0%; Score 678; DB 6; Length 678;
Best Local Similarity 100.0%; Pred. No. 1.2e-189;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGTCCTTCAAGAATGTTATCAAGGATTTCAAGGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 ATGAGGTCCTTCAAGAATGTTATCAAGGATTTCAAGGTTTAAAGTTTCGCATGGAAGGA 60
Qy 61 ACAGTCATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGGAGGCCATACGAAGGC 120
Db 61 ACAGTCATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGGAGGCCATACGAAGGC 120
Qy 121 CACAATACCGTAAAGCTTAAAGGTAACCAAGGGGAGCCCTTGGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAAGGTAACCAAGGGGAGCCCTTGGCTTGGGATATT 180

181 TTGTACACCAATTTTCTAGTATGGAAGCAAGGTATATATGTCAGACACCTGTCGCGACATACCA 240
181 TTGTACACCAATTTTCTAGTATGGAAGCAAGGTATATATGTCAGACACCTGTCGCGACATACCA 240
241 GACTATATAAAGCTGTCTATTTCTGGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
241 GACTATATAAAGCTGTCTATTTCTGGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
361 AAGGTCAAGTTTCATTGGCGTGAACCTTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
361 AAGGTCAAGTTTCATTGGCGTGAACCTTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATGCGGTGTTGAAAGGAGAG 480
421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATGCGGTGTTGAAAGGAGAG 480
481 ATTCAATGAAGCTCTGAAGCTGAAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
481 ATTCAATGAAGCTCTGAAGCTGAAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
541 TACATGCAAGAGAGCTGTGCGACTACCGGTACTACTATGTTGACTCCAACTGGAT 600
541 TACATGCAAGAGAGCTGTGCGACTACCGGTACTACTATGTTGACTCCAACTGGAT 600
601 ATACCAAGCCACCAAGCACTATACAATCTGTCAGCAGTATGAAGAACCGAGGGACGC 660
601 ATACCAAGCCACCAAGCACTATACAATCTGTCAGCAGTATGAAGAACCGAGGGACGC 660
661 CACCATCTGTTCTCTTTAA 678
661 CACCATCTGTTCTCTTTAA 678

RESULT 4
AX233581
LOCUS AX233581 678 bp DNA linear PAT 11-SEP-2001
DEFINITION Sequence 5 from Patent WO0162919.
ACCESSION AX233581
VERSION AX233581.1 GI:15593305

KEYWORDS
SOURCE Discosoma sp.
ORGANISM Discosoma sp.
REFERENCE 1
AUTHORS Nelson, D., Zamaira, E. and Tsien, R.
TITLE Modified fluorescent proteins
JOURNAL Patent: WO 0162919-A 5 30-AUG-2001;
Aurora Biosciences Corporation (US)

FEATURES
source
1..678
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"
/note="red"

ORIGIN
Query Match 100.0%; Score 678; DB 6; Length 678;
Best Local Similarity 100.0%; Pred. No. 1.2e-189;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGGCTTCCAGATGTTATCAAGAGTTCATGAGTTTAAGTTTCGATGAGGA 60
1 ATGAGGCTTCCAGATGTTATCAAGAGTTCATGAGTTTAAGTTTCGATGAGGA 60
61 ACGGTCAATGGCAGCGAGTTTGAATAGAGGCGAAGGAGGGGAGCCATACGAAGGC 120
61 ACGGTCAATGGCAGCGAGTTTGAATAGAGGCGAAGGAGGGGAGCCATACGAAGGC 120

QY 121 CACAATACCGTTAAAGCTTAAAGGTAAACCAAGGGGGACCTTTGGCATTTGCTTGGGATATT 180
Db 121 CACAATACCGTTAAAGCTTAAAGGTAAACCAAGGGGGACCTTTGGCATTTGCTTGGGATATT 180
QY 181 TTGTACACCAATTTTCTAGTATGGAAGCAAGGTATATATGTCAGACACCTGTCGCGACATACCA 240
Db 181 TTGTACACCAATTTTCTAGTATGGAAGCAAGGTATATATGTCAGACACCTGTCGCGACATACCA 240
QY 241 GACTATATAAAGCTGTCTATTTCTGGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATATAAAGCTGTCTATTTCTGGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
QY 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
QY 361 AAGGTCAAGTTTCATTGGCGTGAACCTTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCAAGTTTCATTGGCGTGAACCTTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATGCGGTGTTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATGCGGTGTTGAAAGGAGAG 480
QY 481 ATTCAATGAAGCTCTGAAGCTGAAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Db 481 ATTCAATGAAGCTCTGAAGCTGAAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
QY 541 TACATGCAAGAGAGCTGTGCGACTACCGGTACTACTATGTTGACTCCAACTGGAT 600
Db 541 TACATGCAAGAGAGCTGTGCGACTACCGGTACTACTATGTTGACTCCAACTGGAT 600
QY 601 ATACCAAGCCACCAAGCACTATACAATCTGTCAGCAGTATGAAGAACCGAGGGACGC 660
Db 601 ATACCAAGCCACCAAGCACTATACAATCTGTCAGCAGTATGAAGAACCGAGGGACGC 660
QY 661 CACCATCTGTTCTCTTTAA 678
Db 661 CACCATCTGTTCTCTTTAA 678

RESULT 5
AF168419
LOCUS AF168419 859 bp mRNA linear INV 27-JUL-2001
DEFINITION Discosoma sp. fluorescent protein FP583 mRNA, complete cds.
ACCESSION AF168419
VERSION AF168419.2 GI:7105733

KEYWORDS
SOURCE Discosoma sp.
ORGANISM Discosoma sp.
REFERENCE 1
AUTHORS Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.

REFERENCE 1 (bases 1 to 859)
AUTHORS Matz, M.V., Pradkov, A.F., Labas, Y.A., Savitsky, A.P., Zarskiy, A.G.,
Markelov, M.L. and Lukyanov, S.A.
TITLE Fluorescent proteins from nonbioluminescent Anthozoa species
JOURNAL Nat. Biotechnol. 17 (10), 969-973 (1999)
MEDLINE 99436614
PUBMED 10504696

REFERENCE 2 (bases 1 to 859)
AUTHORS Matz, M.V., Pradkov, A.F., Labas, Y.A., Savitsky, A.P., Zarskiy, A.G.,
Markelov, M.L. and Lukyanov, S.A.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1999) Institute of Bioorganic Chemistry RAS,
Miklukho-Maklaya 16/10, Moscow 117871, Russia

REFERENCE 3 (bases 1 to 859)
AUTHORS Matz, M.V., Pradkov, A.F., Labas, Y.A., Savitsky, A.P., Zarskiy, A.G.,
Markelov, M.L. and Lukyanov, S.A.
TITLE Direct Submission
JOURNAL Submitted (25-FEB-2000) Institute of Bioorganic Chemistry RAS,
Miklukho-Maklaya 16/10, Moscow 117871, Russia
REMARK Sequence update by submitter
COMMENT On Feb 25, 2000 this sequence version replaced gi:6090866.


```
Db 594 TACATGCCAAGAGCCTGTGACGCTACCAAGGTTACTATGTTGACTCCAAACTGGAT 653
Qy 601 ATACCAAGCCACCAAGAGCTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 660
Db 654 ATACCAAGCCACCAAGAGCTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 713
Qy 661 CACATCTGTTCTTTAA 678
Db 714 CACATCTGTTCTTTAA 731

RESULT 7
AF545828 AF545828 678 bp mRNA linear INV 04-JUN-2004
LOCUS Discosoma sp. JW-2002 orange fluorescent protein FP586 mRNA,
DEFINITION complete cds.
ACCESSION AF545828
VERSION AF545828.1 GI:33333763
SOURCE Discosoma sp. JW-2002
ORGANISM Discosoma sp. JW-2002
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
1 (bases 1 to 678)
Wiedenmann, J. and Girod, A.
Direct Submission
Submitted (12-SEP-2002) Abteilung Allgemeine Zoologie und
Endokrinologie, Universitaet Ulm, Albert Einstein Allee 11, Ulm
89069, Germany
LOCATION/Qualifiers
source 1. .678
/organism="Discosoma sp. JW-2002"
/mol_type="mRNA"
/db_xref="taxon:208461"
/notes="isolated from specimen with smooth disc"
CDS 1. .678
/notes="GFP-like protein, orange-red; dsFP586; similar to
fluorescent protein drFP583 (dsRed)"
/codon_start=1
/product="orange fluorescent protein FP586"
/protein_id="AAQ11987.1"
/db_xref="GI:33333764"
/translators="M5CSKNVKEFMFKVPMRKYRMEGTGNGHFEIEGEGRPYGHNTV
KLVTGKPPFPFADILSPQYQSVKVPKADIPDKLSPPEGFKWRVWVNFEDG
GVTVTQDSLODGCFTYKVFIGNPSPDPVMOKTGMGWEASTERLYPRDGVLGKD
IHKALKKDGSHVLEPKSIYMAKPKVQLPFGYVYVDSKLDITSHNEDYTIQEYRTE
GRHHLFL"

ORIGIN
Query Match 98.3%; Score 666.8; DB 3; Length 678;
Best Local Similarity 99.0%; Pred. No. 2.5e-186;
Matches 671; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ATGAGGTCTTCCAGAAATGTTATCAAGAGTTTCATAGGTTTAAGTTTCGCATGGAAGGA 60
Db 1 ATGAGTTGTTTCCAGAAATGTTATCAAGAGTTTCATAGGTTTAAGTTTCGCATGGAAGGA 60
Qy 61 ACGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGGAGGAGGCCATACGAGGC 120
Db 61 ACGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGGAGGAGGCCATACGAGGC 120
Qy 121 CACAATACCGTAAAGCTTAAGTAAACCAAGGGGACCTTTGCCATTGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAGTAAACCAAGGGGACCTTTGCCATTGCTTGGGATATT 180
Qy 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCACGACCCCTGCCGACATACCA 240
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCACGACCCCTGCCGACATACCA 240
Qy 241 GACTATAAAAGCTGTGCTATTTCTGGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATAAAAGCTGTGCTATTTCTGGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
```

```
Qy 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGGTCAAGTTCAATTCGGCGTGAACCTTTCCTTCGATGGACCTGTGTATGCAAAAGAACACA 420
Db 361 AAGGTCAAGTTCAATTCGGCGTGAACCTTTCCTTCGATGGACCTGTGTATGCAAAAGAACACA 420
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGTTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGTTGAAAGGAGAT 480
Qy 481 ATTCAATAGGCTCTGAAGCTGAAAGAGCGTGTTCATCTACCTAGTTGAAATTCAAAGATTT 540
Db 481 ATTCAATAGGCTCTGAAGCTGAAAGAGCGTGTTCATCTACCTAGTTGAAATTCAAAGATTT 540
Qy 541 TACATGGCAAAAGAGCGTGTGACGCTACCAAGGCTACTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAAAGAGCGTGTGACGCTACCAAGGCTACTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATAAACAGCCACCAACGAACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 660
Db 601 ATAAACAGCCACCAACGAACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 660
Qy 661 CACATCTGTTCTTTAA 678
Db 661 CACATCTGTTCTTTAA 678

RESULT 8
AX348043 AX348043 666 bp DNA linear PAT 06-FEB-2002
LOCUS Sequence 22 from Patent EP1172375.
DEFINITION AX348043
ACCESSION AX348043
VERSION AX348043.1 GI:18614153
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Odenthal, M. and Jung, D.
TITLE Gene expression, genome alteration and reporter gene expression in myofibroblasts and myofibroblasts-like cells by using the regulatory regions within the alpha smooth muscle alpha-actin gene Patent: EP 1172375-A 22 16-JAN-2002;
JOURNAL Odenthal, Margarete (DE)
FEATURES Location/Qualifiers
source 1. .666
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Red Fluorescent Protein"

ORIGIN
Query Match 98.1%; Score 665; DB 6; Length 666;
Best Local Similarity 100.0%; Pred. No. 8.4e-186;
Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 AAGAATGTTATCAAGGAGTTTCATAGGTTTAAGTTTCGCATGAGGAAACGTTCAATGGG 72
Db 1 AAGAATGTTATCAAGGAGTTTCATAGGTTTAAGTTTCGCATGAGGAAACGTTCAATGGG 60
Qy 73 CACGAGTTTCAATAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 132
Db 61 CACGAGTTTCAATAGAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 120
Qy 133 AAGCTTAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCAACCAAA 192
Db 121 AAGCTTAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCAACCAAA 180
Qy 193 TTTCAGTATGGAACGAAGGTATATGTCAAGCACCTGCCGACATACACAGACTATAAAG 252
Db 193 TTTCAGTATGGAACGAAGGTATATGTCAAGCACCTGCCGACATACACAGACTATAAAG 252
```

```
Db 181 TTTAGTATGAAGCAAGGTATATGTCAAGCACCTCGCGACATACCAGACTATAAAG 240
Qy 253 CTGTCAATTCCTGAAGGATTTAAATGGAAAGGTCATGAACCTTTGAAGCGGTGGGTC 312
Db 241 CTGTCAATTCCTGAAGGATTTAAATGGAAAGGTCATGAACCTTTGAAGCGGTGGGTC 300
Qy 313 GTTACTCTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTACAAGGTCAGTTC 372
Db 301 GTTACTCTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTACAAGGTCAGTTC 360
Qy 373 ATTGGCGTGAACCTTCCTCGATGGACCTGTTATGCAAAAGAGACAATGGGCTGGAA 432
Db 361 ATTGGCGTGAACCTTCCTCGATGGACCTGTTATGCAAAAGAGACAATGGGCTGGAA 420
Qy 433 GCAGCACTGAGCGTTTGTATCCTCGTGGTGGCTGTTGAAAGGAGAGATTCTAAGGCT 492
Db 421 GCCAGCACTGAGCGTTTGTATCCTCGTGGTGGCTGTTGAAAGGAGAGATTCTAAGGCT 480
Qy 493 CTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAAATTCAAAAGTATTACATGGCAAG 552
Db 481 CTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAAATTCAAAAGTATTACATGGCAAG 540
Qy 553 AGCCTGTGAGCTACAGGTTACTATGTTGACTCCAAAACCTGGATATAACAAGCCAC 612
Db 541 AGCCTGTGAGCTACAGGTTACTATGTTGACTCCAAAACCTGGATATAACAAGCCAC 600
Qy 613 AACGAAGACTATCAATCGTTGAGCAGTATGAAGAACCGAGGACCGCCACTCTGTC 672
Db 601 AACGAAGACTATCAATCGTTGAGCAGTATGAAGAACCGAGGACCGCCACTCTGTC 660
Qy 673 CTTTA 677
Db 661 CTTTA 665

RESULT 9
AX353910
LOCUS AX353910 666 bp DNA linear PAT 06-FEB-2002
DEFINITION Sequence 22 from Patent WO0204509.
ACCESSION AX353910
VERSION AX353910.1 GI:18618888
KEYWORDS synthetic construct
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS Odenthal,M. and Jung,D.
TITLE Gene expression, genome alteration and reporter expression in myofibroblasts and myofibroblast-like cells
JOURNAL Patent: WO 0204509-A 22 17-JAN-2002;
Odenthal, Margarete (DE)
FEATURES
    source
        location/Qualifiers
            1..666
                /organism="synthetic construct"
                /mol_type="unassigned DNA"
                /db_xref="taxon:32630"
                /note="Red Fluorescent Protein"
ORIGIN
Query Match 98.1%; Score 665; DB 6; Length 666;
Best Local Similarity 100.0%; Pred. No. 8.4e-186;
Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 AAGATGTTTCAAGGAGTTTCATGAGTTTAAAGTTTCGATGGAAGAACGGTCAATGGG 72
Db 1 AAGATGTTTCAAGGAGTTTCATGAGTTTAAAGTTTCGATGGAAGAACGGTCAATGGG 60
Qy 73 CACGAGTTTGAATAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 132
Db 61 CACGAGTTTGAATAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
Qy 133 AAGCTTAAGGTAACCAAGGGGGGACCTTTCGCAATTCCTTGGGATATTTTGTCAACCAA 192
```

```
Db 121 AAGCTTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCAACCAA 180
Qy 193 TTTTCAGTATGGAAGCAAGGTATATGTCACAGCACCTCGCGACATACCAGACTATAAAG 252
Db 181 TTTTCAGTATGGAAGCAAGGTATATGTCACAGCACCTCGCGACATACCAGACTATAAAG 240
Qy 253 CTGTCAATTCCTGAAGGATTTAAATGGAAAGGTCATGAACCTTTGAAGCGGTGGGTC 312
Db 241 CTGTCAATTCCTGAAGGATTTAAATGGAAAGGTCATGAACCTTTGAAGCGGTGGGTC 300
Qy 313 GTTACTCTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTACAAGGTCAGTTC 372
Db 301 GTTACTCTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTACAAGGTCAGTTC 360
Qy 373 ATTGGCGTGAACCTTCCTCGATGGACCTGTTATGCAAAAGAGACAATGGGCTGGAA 432
Db 361 ATTGGCGTGAACCTTCCTCGATGGACCTGTTATGCAAAAGAGACAATGGGCTGGAA 420
Qy 433 GCAGCACTGAGCGTTTGTATCCTCGTGGTGGCTGTTGAAAGGAGAGATTCTAAGGCT 492
Db 421 GCCAGCACTGAGCGTTTGTATCCTCGTGGTGGCTGTTGAAAGGAGAGATTCTAAGGCT 480
Qy 493 CTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAAATTCAAAAGTATTACATGGCAAG 552
Db 481 CTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAAATTCAAAAGTATTACATGGCAAG 540
Qy 553 AGCCTGTGAGCTACAGGTTACTATGTTGACTCCAAAACCTGGATATAACAAGCCAC 612
Db 541 AGCCTGTGAGCTACAGGTTACTATGTTGACTCCAAAACCTGGATATAACAAGCCAC 600
Qy 613 AACGAAGACTATCAATCGTTGAGCAGTATGAAGAACCGAGGACCGCCACTCTGTC 672
Db 601 AACGAAGACTATCAATCGTTGAGCAGTATGAAGAACCGAGGACCGCCACTCTGTC 660
Qy 673 CTTTA 677
Db 661 CTTTA 665

RESULT 10
AY679107
LOCUS AY679107 711 bp mRNA linear INV 25-AUG-2004
DEFINITION Discosoma sp. RC-2004 enhanced red fluorescent protein R+ mRNA, complete cds.
ACCESSION AY679107
VERSION AY679107.1 GI:51472046
KEYWORDS Discosoma sp. RC-2004
SOURCE Discosoma sp. RC-2004
ORGANISM Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE 1 (bases 1 to 711)
AUTHORS Carter,R.W., Gibbs,P.D.L. and Schmale,M.C.
TITLE Cloning of Anthozoan Fluorescent Protein Genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 711)
AUTHORS Carter,R.W., Gibbs,P.D.L. and Schmale,M.C.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2004) MBF, University of Miami, 4600 Rickenbacker Cswy, Miami, FL 33129, USA
FEATURES
    source
        location/Qualifiers
            1..711
                /organism="Discosoma sp. RC-2004"
                /mol_type="mRNA"
                /db_xref="taxon:289055"
            1..711
                /note="GFP-like; DsRed+"
                /codon_start=1
                /product="enhanced red fluorescent protein R+"
                /protein_id="AAU04444.1"
                /db_xref="GI:51472047"
                /translation="MSSCKNVIKEFMRFKVRMEGTVNGHFEIIEGEGRPFYEGHNTV
                KLVKTKGGLPFAWDILSPQFYQSKVYVVKHPADIPTYKLSFPGEFKWVNMFPDG
```


GVVTVTDSSLDQSCFIYKVKFIGVNPSPDPVMQKTMGWEASTERLYPRDGVKXGE
IHKALKLKDGHVLYEFKTYIMAKKPVQLPGYVYVDSKLDITSHNKDYTIIVEQYERTE
GRHLLFLKSELGSNVGER"

ORIGIN

Query Match 97.5%; Score 661; DB 3; Length 711;
Best Local Similarity 98.5%; Pred. No. 1.3e-184;
Matches 667; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 ATGAGGTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 60
DB 1 ATGAGTGTGTTCCAAGATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGTATGGAAGGA 60
QY 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 120
DB 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 120
QY 121 CACAATACCGTAAAGCTTAAAGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB 121 CACAATACCGTAAAGCTTAAAGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
QY 181 TTGTACCAACAATTTCAATGAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 240
DB 181 TTGTACCAACAATTTCAATGAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 240
QY 241 GACTATAAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB 241 GACTATAAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
QY 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGCGTTTCATCTAC 360
DB 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGCGTTTCATCTAC 360
QY 361 AAGTCAAGTTCATTGGCGTGAACCTTTCTTCGATGAGACCTGTTATGCAAGAGACA 420
DB 361 AAGTCAAGTTCATTGGCGTGAACCTTTCTTCGATGAGACCTGTTATGCAAGAGACA 420
QY 421 ATGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCGTGAAGCGGTTGTTGAAGAGAG 480
DB 421 ATGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCGTGAAGCGGTTGTTGAAGAGAG 480
QY 481 ATTCAATAGGCTCGAAGCTGGAAGCGGTTGTTTACCTAGTTGAATTCGAAGGATTT 540
DB 481 ATTCAATAGGCTCGAAGCTGGAAGCGGTTGTTTACCTAGTTGAATTCGAAGGATTT 540
QY 541 TACATGCAAGAGAGCCTGTGCAGCTACCAAGGTTACTATGTTGACTCCAAACTGGAT 600
DB 541 TACATGCAAGAGAGCCTGTGCAGCTACCAAGGTTACTATGTTGACTCCAAACTGGAT 600
QY 601 ATAAAGCCACAAAGAGCTATACAACTGTTGAGCAGTATGAAGAACCGAGGGACGC 660
DB 601 ATAAAGCCACAAAGAGCTATACAACTGTTGAGCAGTATGAAGAACCGAGGGACGC 660
QY 661 CACCATCTGTTCTTTAA 677
DB 661 CACCATCTGTTCTTTAA 677

RESULT 11

AX686888
LOCUS AX686888 898 bp DNA linear PAT 31-MAR-2003
DEFINITION Sequence 11 from Patent WO0127150.
ACCESSION AX686888
VERSION AX686888.1 GI:29409468
KEYWORDS
SOURCE Discosoma sp.
ORGANISM Discosoma sp.

REFERENCE 1
AUTHORS Lukyanov, S.A., Pradkov, A.P., Labas, Y.A., Matz, M.V. and Teresikh, A.
TITLE Anthozoa derived chromo/fluoroproteins and methods for using the same

JOURNAL Patent: WO 0127150-A 11 19-APR-2001;
Clontech Laboratories Inc. (US)
FEATURES Location/Qualifiers
source 1..898
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"

ORIGIN

Query Match 97.2%; Score 658.8; DB 6; Length 898;
Best Local Similarity 98.2%; Pred. No. 5.9e-184;
Matches 666; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 ATGAGGTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 60
DB 93 ATGAGGTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 152
QY 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 120
DB 153 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 212
QY 121 CACAATAACCGTAAAGCTTAAAGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB 213 CACAATAACCGTAAAGCTTAAAGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 272
QY 181 TTGTACCAACAATTTCAATGAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 240
DB 273 TTGTACCAACAATTTCAATGAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 332
QY 241 GACTATAAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB 333 GACTATAAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 392
QY 301 GACGGTGGCGTCGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGCGTTTTCATCTAC 360
DB 393 GACGGTGGCGTCGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGCGTTTTCATCTAC 452
QY 361 AAGTCAAGTTCATTGGCGTGAACCTTTCTTCGATGAGACCTGTTATGCAAGAGACA 420
DB 453 AAGTCAAGTTCATTGGCGTGAACCTTTCTTCGATGAGACCTGTTATGCAAGAGACA 512
QY 421 ATGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCGTGAAGCGGTTGTTGAAGAGAG 480
DB 513 ATGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCGTGAAGCGGTTGTTGAAGAGAG 572
QY 481 ATTCAATAGGCTCGAAGCTGGAAGCGGTTGTTTACCTAGTTGAATTCGAAGGATTT 540
DB 573 ATTCAATAGGCTCGAAGCTGGAAGCGGTTGTTTACCTAGTTGAATTCGAAGGATTT 632
QY 541 TACATGCAAGAGAGCCTGTGCAGCTACCAAGGTTACTATGTTGACTCCAAACTGGAT 600
DB 633 TACATGCAAGAGAGCCTGTGCAGCTACCAAGGTTACTATGTTGACTCCAAACTGGAT 692
QY 601 ATAAAGCCACAAAGAGCTATACAACTGTTGAGCAGTATGAAGAACCGAGGGACGC 660
DB 693 ATAAAGCCACAAAGAGCTATACAACTGTTGAGCAGTATGAAGAACCGAGGGACGC 752
QY 661 CACCATCTGTTCTTTAA 678
DB 753 CACCATCTGTTCTTTAA 770

RESULT 12

AX679106
LOCUS AX679106 921 bp mRNA linear INV 25-AUG-2004
DEFINITION Discosoma sp. RC-2004 red fluorescent protein R1 mRNA, complete cds.
ACCESSION AY679106
VERSION AY679106.1 GI:51472044
KEYWORDS
SOURCE Discosoma sp. RC-2004
ORGANISM Discosoma sp. RC-2004
Eukaryota; Metazoa; Anthozoa; Cnidaria; Anthozoa; Zoantharia;

Corallimorpharia; Discosomatidae; Discosoma.
1 (bases 1 to 921)
Carter, R.W., Gibbs, P.D.L. and Schmale, M.C.
Cloning of Anthozoan Fluorescent Protein Genes
Unpublished
2 (bases 1 to 921)
Carter, R.W., Gibbs, P.D.L. and Schmale, M.C.
Direct Submission
Submitted (06-JUL-2004) MBF, University of Miami, 4600 Rickenbacker
Cswy, Miami, FL 33129, USA
Location/Qualifiers
FEATURES
source
1..921
/organism="Discosoma sp. RC-2004"
/mol_type="mRNA"
/db_xref="taxon:289055"
1..66
67..777
/note="GFP-like; DepR1"
/codon_start=1
/product="red fluorescent protein R1"
/protein_id="AAU04443.1"
/db_xref="GI:51472045"
/translation="MCSKNVKEPMRFKVRMEGTVNGHEFEIGEGEPYRGHTV
KLKVTGGPFLAWDLISLPQFYQYKSVYVHPADIPDYKLSFPEGFKVRVNMFDG
GVVTYQDPSLPQCFYKVFIGNPSPDGVNQKTMGWEASTERLYPRDGLVGE
IHKALKLDGGHYLVFEFTIYMAKVPQLPGYVYVDSKLDITSHNKDYTIVEQYRTE
GRHHLFKAEGLSNVGR"
778..921
3'UTR
polya_signal
882..887
ORIGIN
Query Match 96.8%; Score 656.2; DB 3; Length 921;
Best Local Similarity 98.1%; Pred. No. 3.5e-183;
Matches 664; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 1 ATGAGTGCTTCCAAGATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATGAAGGA 60
Db 67 ATGAGTTGTTTCCAAGATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATGAAGGA 126
Qy 61 ACGGTCAATGGCAGGATTTGAATAGAGGCGAAGGAGGAGGAGCCATACGAGGC 120
Db 127 ACGGTCAATGGCAGGATTTGAATAGAGGCGAAGGAGGAGGAGCCATACGAGGC 186
Qy 121 CACAATACCGTAAAGCTTAAAGTTAAACCAAGGGGACCTTTGCCATTTGCTTGGATATT 180
Db 187 CACAATACCGTAAAGCTTAAAGTTAAACCAAGGGGACCTTTGCCATTTGCTTGGATATT 246
Qy 181 TTGTACCAACAATTTCAAGTATGGAAGCAAGGATATATGTCAGCAACCTCGCCGACATACCA 240
Db 247 TTGTACCAACAATTTCAAGTATGGAAGCAAGGATATATGTCAGCAACCTCGCCGACATACCA 306
Qy 241 GACTATAAAGCTGTCATTTCTGTAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 307 GACTATAAAGCTGTCATTTCTGTAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 366
Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCAAGTTTGAGGATGCTGTTTCATCTAC 360
Db 367 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCAAGTTTGAGGATGCTGTTTCATCTAC 426
Qy 361 AAGGTCAAGTTCATGTCGGTGAACTTTCTTCGATCGAAGCTGTTATGCAAAAAGAGACA 420
Db 427 AAGGTCAAGTTCATGTCGGTGAACTTTCTTCGATCGAAGCTGTTATGCAAAAAGAGACA 486
Qy 421 ATGGGCTGGGAAGCCAGCAGCTGAGCGTTTGATCTCGTATGCGCGTGTGAAGGAGAG 480
Db 487 ATGGGCTGGGAAGCCAGCAGCTGAGCGTTTGATCTCGTATGCGCGTGTGAAGGAGAG 546
Qy 481 ATTCAAGGCTCTGAAGCTGAAGAGCGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 540
Db 547 ATTCAAGGCTCTGAAGTGAAGAGCGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 606
Qy 541 TACATGCAAAAGACCTGTGACGCTACGAGGTAACGAGGTAATGTTGATCTCAAACTGGAT 600

Db 607 TACATGCAAAAGACCTGTGACGCTACCGAGGTACTACTATGTTGACTTCAAACTGGAT 666
Qy 601 ATAAACAGCCACAAACGAGACTATACAATCGTTTGACAGTATCAAAAGAACCGAGGACGC 660
Db 667 ATAAACAGCCACAAACGAGACTATACAATCGTTTGACAGTATCAAAAGAACCGAGGACGC 726
Qy 661 CACCATCTGTTCCTTTA 677
Db 727 CACCATCTGTTCCTTAA 743
RESULT 13
AF272711
LOCUS
DEFINITION
876 bp mRNA linear INV 26-SEP-2000
Discosoma sp. SSAL-2000 red fluorescent protein (FP593) mRNA,
complete cds.
AF272711 GI:10304306
ORGANISM
Discosoma sp. SSAL-2000
Discosoma sp. SSAL-2000
Eukaryota; Metazoa; Chnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE
AUTHORS
1 (bases 1 to 876)
Pradkov, A.F., Chen, Y., Ding, L., Barsova, E.V., Matz, M.V. and
Lukyanov, S.A.
TITLE
Novel fluorescent protein from Discosoma coral and its mutants
possesses a unique far-red fluorescence
FEBS Lett. 479 (3), 127-130 (2000)
20434599
10981720
REFERENCE
AUTHORS
2 (bases 1 to 876)
Pradkov, A.F., Chen, Y., Ding, L., Barsova, E.V., Matz, M.V. and
Lukyanov, S.A.
TITLE
Direct Submission
JOURNAL
Submitted (26-MAY-2000) Institute of Bioorganic Chemistry RAS,
Miklukho-Maklaya 16/10, Moscow 117871, Russia
FEATURES
Location/Qualifiers
1..876
/organism="Discosoma sp. SSAL-2000"
/mol_type="mRNA"
/db_xref="taxon:137428"
1..876
/genes="FP593"
45..737
/gene="FP593"
/note="dsFP593"
/codon_start=1
/product="red fluorescent protein"
/protein_id="AAG16224.1"
/db_xref="GI:10304307"
/translation="MCSKNVKEPMRFKVRMEGTVNGHEFEIKGEGRPFVGHCSV
KLKVTGGPFLAFDILSPQFYQYKSVYVHPADIPDYKLSFPEGFKVRVNMFDG
GVVTYQDPSLPQCFYKVFIGNPSPDGVNQKTMGWEASTERLYPRDGLVGE
IHKALKLDGGHYLVFEFTIYMAKVPQLPGYVYVDSKLDITSHNKDYTIVEQYRTE
GRHHLFPKPLQ"
QGRHHPFKPLQ"
ORIGIN
Query Match 83.5%; Score 566.4; DB 3; Length 876;
Best Local Similarity 90.6%; Pred. No. 1.5e-156;
Matches 616; Conservative 0; Mismatches 61; Indels 3; Gaps 1;
Qy 1 ATGAGTGCTTCCAAGATGTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGATGAAGGA 60
Db 45 ATGAGTTGTTCCAAAGATGTGATCAAGGAGTTTCATGAGTTTCGATGAAGGA 104
Qy 61 ACGGTCAATGGCAGCAGTTTGAATAGAGCGCAAGGAGGAGGAGCCATACGAGGC 120
Db 105 ACGGTCAATGGCAGCAGTTTGAATAGAGCGCAAGGAGGAGGAGCCATACGAGGT 164
Qy 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGATATT 180
Db 165 CACTGTTCCGTAAGAGCTTATGGTAAACCAAGGGTGGACCTTTGCCATTTGCTTTGATATT 224

181 TTGTCCACACAAATTTCTAGTATGGAAGCAAGGTATATGTCTCAGACACCTTGCAGCATACCA 240
182 |||||
225 TTGTCCACACAAATTTCTAGTATGGAAGCAAGGTATATGTCTCAGACACCTTGCAGCATACCA 284
226 |||||
241 GACTATATAAAGCTGTCTCATTTCTCTGAAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
242 |||||
285 GACTATATAAAGCTGTCTCATTTCTCTGAAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 344
286 |||||
301 GACGGTGGCGTCTGTTACTGTATACCCAGGATTTCCAGTTTTCAGAGTGGCTGTTTCATCTAC 360
302 |||||
345 GACGGTGGCGTCTGTTACTGTATACCCAGGATTTCCAGTTTTCAGAGTGGCTGTTTCATCTAC 404
346 |||||
361 AAGGTCAAGTTTCATTGGCGTGAATCTTCTTCCAGTGGACCTGTATGCAAAAGAGACA 420
362 |||||
405 GAGGTCAAGTTTCATTGGCGTGAATCTTCTTCTGATGGACCTGTATGCAAAAGAGACA 464
406 |||||
421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGTAAGAGAGAG 480
422 |||||
465 CGGGGCTGGGAAGCCAGCTCTGAGCGTTTGTATCTCTGATGGCGTGTGTAAGAGAGAG 524
466 |||||
481 ATTCAATAGGCTCTGAAGCTGGAAGCGGTGCTCATTTACCTAGTTTGAATTCAAAAGTATT 540
482 |||||
525 ATCCATATGGCTCTGAGGCTGGAAGGAGCGGCCATTACCTCGTTGAATTCAAAAGTATT 584
526 |||||
541 TACATGCAAGAGAGCGCT--GTGACGCTACCGGTTACTATGTTGACTCCAAACTG 597
542 |||||
585 TACATGTAAGAAGCGCTTCAGTGCAGTGGCGGCTTACTATGTTGACTCCAAACTG 644
586 |||||
598 GATATAACAAGCCACACAGCACTATACAAATCGTTGAGCAGTATGAAAGACCGAGGA 657
599 |||||
645 GATATGACGAGCCACACAGCACTATACAAATCGTTGAGCAGTATGAAAGACCGAGGA 704
646 |||||
658 CGCCACCCTCTGTTCTTTA 677
705 CGCCACCCTCTGTTCTTTA 724

RESULT 14
AX686894
LOCUS
DEFINITION
Sequence 17 from Patent WO0127150.
ACCESSION
AX686894
VERSION
AX686894.1 GI:29409472
KEYWORDS
Discosoma sp.
SOURCE
ORGANISM
Discosoma sp.
Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia;
Corallimorpharia; Discosomatidae; Discosoma.
REFERENCE
1
AUTHORS
Lukyanov,S.A., Pradkov,A.F., Labas,Y.A., Matz,M.V. and Terskikh,A.
TITLE
Anthozoa derived chromo/fluoroproteins and methods for using the
JOURNAL
same
Patent: WO 0127150-A 17 19-APR-2001;
Clontech Laboratories Inc. (US)
FEATURES
source
1. .876
/organism="Discosoma sp."
/mol_type="unassigned DNA"
/db_xref="taxon:86600"

ORIGIN

Query Match 83.5%; Score 566.4; DB 6; Length 876;
Best Local Similarity 90.6%; Pred. No. 1.5e-156;
Matches 616; Conservative 0; Mismatches 61; Indels 3; Gaps 1;
QY 1 ATGAGGTCTTCCAGAAATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATGGAAGGA 60
DB 45 ATGAGTTGTTTCCAAAGAAATGTATCAAGAGTTTCATGAGTTTCAGGTTTCGATGGAAGGA 104
QY 61 ACAGTCAATGGGACGAGTTTGAATAGAAGCGGCAAGAGGAGGAGGAGGCCATACGAAGGC 120
DB 105 ACAGTCAATGGGACGAGTTTGAATAGAAGCGGCAAGAGGAGGAGGAGGCCCTTACGAAGGT 164

QY 121 CACATATACCGTAAGCTTAAGGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB 165 CACTGTTCCGTAAGCTTAAGGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 224
QY 181 TTGTCCACACAAATTTCTAGTATGGAAGCAAGGTATATGTCTCAGACACCTTGCAGCATACCA 240
DB 225 TTGTCCACACAAATTTCTAGTATGGAAGCAAGGTATATGTCTCAGACACCTTGCAGCATACCA 284
QY 241 GACTATATAAAGCTGTCTCATTTCTCTGAAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
DB 285 GACTATATAAAGCTGTCTCATTTCTCTGAAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 344
QY 301 GACGGTGGCGTCTGTTACTGTATACCCAGGATTTCCAGTTTTCAGAGTGGCTGTTTCATCTAC 360
DB 345 GACGGTGGCGTCTGTTACTGTATACCCAGGATTTCCAGTTTTCAGAGTGGCTGTTTCATCTAC 404
QY 361 AAGGTCAAGTTTCATTGGCGTGAATCTTCTTCCAGTGGACCTGTATGCAAAAGAGACA 420
DB 405 GAGGTCAAGTTTCATTGGCGTGAATCTTCTTCTGATGGACCTGTATGCAAAAGAGACA 464
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGTAAGAGAGAG 480
DB 465 CGGGGCTGGGAAGCCAGCTCTGAGCGTTTGTATCTCTGATGGCGTGTGTAAGAGAGAG 524
QY 481 ATTCAATAGGCTCTGAAGCTGGAAGCGGTGCTCATTTACCTAGTTTGAATTCAAAAGTATT 540
DB 525 ATCCATATGGCTCTGAGGCTGGAAGGAGCGGCCATTACCTCGTTGAATTCAAAAGTATT 584
QY 541 TACATGCAAGAGAGCGCT--GTGACGCTACCGGTTACTATGTTGACTCCAAACTG 597
DB 585 TACATGTAAGAAGCGCTTCAGTGCAGTGGCGGCTTACTATGTTGACTCCAAACTG 644
QY 598 GATATAACAAGCCACACAGCACTATACAAATCGTTGAGCAGTATGAAAGACCGAGGA 657
DB 645 GATATGACGAGCCACACAGCACTATACAAATCGTTGAGCAGTATGAAAGACCGAGGA 704
QY 658 CGCCACCCTCTGTTCTTTA 677
DB 705 CGCCACCCTCTGTTCTTTA 724

RESULT 15

AR527331
LOCUS
DEFINITION
Sequence 1 from patent US 6723537.
ACCESSION
AR527331
VERSION
AR527331.1 GI:53914309
KEYWORDS
SOURCE
Unknown.
ORGANISM
Unclassified.
REFERENCE
1 (bases 1 to 723)
AUTHORS
Peelle,B.
TITLE
Directed evolution of protein in mammalian cells
JOURNAL
Patent: US 6723537-A 1 20-APR-2004;
FEATURES
Location/Qualifiers
source
1. .723
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 66.5%; Score 450.6; DB 6; Length 723;
Best Local Similarity 79.3%; Pred. No. 3.3e-122;
Matches 534; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 2 TGAGGTCTTCCAGAAATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATGGAAGGA 61
DB 5 TGCGCTCTCTCAAGAAAGCTCATCAAGAGAGTTTCATGCGCTTCAAGGTCGCGATGAGGGA 64
QY 62 CGGTCAATGGGACGAGTTTGAATAGAAGCGGCAAGAGAGGAGGAGGCCATACGAAGGCC 121
DB 65 CCGTGAACGGGCCAGAGTTTCGAGATCGAGGCGGAGGCGGCCCTTACGAGGGCC 124

```
QY 122 ACAATACCGTAAAGCTTAAGTTAACCAAGGGGACCTTTGCCATTTGCTTGGGATATTT 181
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 125 ACAACACCGTGAAGCTGAAGTGACCAAGGGGGGCCCCCTGCCCTTGCCTTGGGACATCC 184
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 TGTCAACACAAATTTAGTATGGAAGCAAGGTATATGTCAAAGCACCCCTGCGGACATACCAG 241
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 185 TGTCCCCCAGTTTCCAGTACGGCTCCAAAGGTGACGTGAAGCACCCCGCGGACATCCCCG 244
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 ACTATAAAGCTGTCTATTTCTTGNAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAAG 301
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 245 ACTACAAGAAGCTGTCTCTTCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACCTTCGAGG 304
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 302 ACGGTGCGCTCGTTACTGTAAACCCAGGATTCACAGTTTGCAGGATGGCTGTTTCATCTACA 361
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 305 ACGGCGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCGAGGAGGCTGCTTCATCTACA 364
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 362 AGGTCAAGTTTCATGGCGTGAACCTTCCCTCCGATGGACCTGTATGCAAAAGAACAA 421
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 365 AGGTGAAGTTTCATCGGCGTGAACCTTCCCTCCGACGGCCCGTAAATGCAGAAGAACCA 424
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 422 TGGGCTGGGAAGCCAGCACCTGAGGGTTTGATCCTCGTGATGGCGTGTGAAAGGAGAGA 481
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 425 TGGGCTGGGAAGGCGCTCCACGAGCGCTGTACCCCGCGAGCGGCTGTGAAGGGCGAGA 484
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 482 TTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCAATTACCTAGTTGAATTCAAAAGTATTT 541
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 485 TCCACAAGGCGCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGAGTATCT 544
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 542 ACATGGCAAGAGCGCTGTGAGCTACAGGGTACTACTATGTTGACTCCAAACTGGATA 601
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 545 ACATGGCCAAGAGCGCGTGCAGCTGCCCGCTACTACTACGTGGACTCCCAAGCTGGACA 604
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 602 TAACAGCCACACGAGACTATACATCGTTGAGCAGTATGAAGAACCGAGGGAGCGCC 661
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 605 TCACCTCCCAACACGAGACTACACCATCTGTGGAGCAGTACGAGCGCACCGAGGGCGGCC 664
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 662 ACCATCTGTTTCCT 674
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 665 ACCACCTGTTTCCT 677
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 1, 2005, 18:20:08
Job time : 3349 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 14:18:00 ; Search time 3040 Seconds
(without alignments)
8489.334 Million cell updates/sec

Title: US-10-006-922a-11

Perfect score: 678

Sequence: 1 atgagggtcttccaagaatgt.....gccaccatctgttcctttaa 678

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: gb_est1.*

2: gb_est2.*

3: gb_hic.*

4: gb_est3.*

5: gb_est4.*

6: gb_est5.*

7: gb_est6.*

8: gb_ges1.*

9: gb_ges2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	7.4	671	4	BI381780
2	49	7.2	681	4	BI386888
3	48.4	7.1	561	4	BI379061
4	48.4	7.1	577	4	BI386955
5	45.8	6.8	459	4	BI386952
6	45.2	6.7	591	4	BI382373
7	44.8	6.6	707	4	BI382442
8	43.6	6.4	537	4	BI376909
9	43.2	6.4	629	4	BI384961
10	41.2	6.1	930	9	CL065858
11	39.8	5.9	472	2	BB854015
12	39.2	5.8	554	4	BI815946
13	39	5.8	592	1	AV719715
14	39	5.8	808	8	BI213346
15	38.6	5.7	571	5	BO451564
16	38.6	5.7	580	5	BO451564
17	38.6	5.7	600	5	BU497757
18	38.6	5.7	604	5	BQ597249
19	38.6	5.7	607	5	BQ739542
20	38.6	5.7	616	5	BQ451730
21	38.4	5.7	571	4	BI387406
22	38.4	5.7	687	4	BI387917
23	38.2	5.6	629	4	BI380008
24	38	5.6	480	9	CE292671

25	37.8	5.6	797	9	CNS003F8
26	37.8	5.6	1025	9	CNS07CE9
27	37.6	5.5	1168	8	CC208512
28	37.2	5.5	548	4	BM521209
29	37.2	5.5	558	6	CD414787
30	37.2	5.5	704	4	BG788189
31	37.2	5.5	888	8	BZ879331
32	36.8	5.4	457	2	BE679790
33	36.8	5.4	536	8	AZ450406
34	36.8	5.4	661	4	BI385436
35	36.6	5.4	453	4	BI385638
36	36.6	5.4	705	2	BE357640
37	36.4	5.4	524	6	CB602899
38	36.4	5.4	558	9	CG117372
39	36.4	5.4	596	6	CB390628
40	36.4	5.4	930	9	CG041582
41	36.2	5.3	725	8	BZ260052
42	36.2	5.3	828	8	BZ248767
43	36	5.3	513	8	BH111996
44	36	5.3	645	9	CE190471
45	36	5.3	679	8	BH188917

ALIGNMENTS

RESULT 1
BI381780
LOCUS
DEFINITION
BFLG1_003411 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPWGp498) Branchiostoma floridae cDNA clone MPWGp498C047 5', mRNA sequence.
671 bp mRNA linear EST 26-AUG-2003
BI381780
BI381780.1 GI:30916875
EST.
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Sukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae; Branchiostoma.
1 (bases 1 to 671)
Panopoulou G., Hennig S., Groth D., Krause A., Poustka A.J., Herwig R., Vingron M. and Lehrach H.
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
Genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GGTATTACCGCAGTCGGGAAGGGGATGTG 3' (M13FSP)
Insert length: 1200 Std Error: 0.00
Seq primer: 5'-CCGTCGCGATTCGCGGT-3' pSport3/86
High quality sequence stop: 671.

JOURNAL
MEDLINE
PUBMED
COMMENT

FEATURES	source	Location/Qualifiers
		1..671
		/organism="Branchiostoma floridae"
		/mol_type="mRNA"
		/db_xref="taxon:7739"
		/clone="MPMGp498C047"
		/tissue_type="whole embryo"
		/dev_stage="5-6 hrs (gastrula stage)"
		/lab_host="E.coli, Xli blue"
		/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
		/note="vector: pSport1; Site_1: SalI, KpnI, EcoRI (5'); Site_2: NotI, BamHI, HindIII (3'); Oligodt primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTCTGACTGACGCGCGCC (7)15-3' and a SalI 5'-TCGACCCACGCGCTCCG-3' adapters (Gibco BRL)."
ORIGIN		
	Query Match	7.4%; Score 50; DB 4; Length 671;
	Best Local Similarity	48.1%; Pred. No. 0.0011;
	Matches	174; Conservative 0; Mismatches 185; Indels 3; Gaps 1;
QY	47	TTCCGATGGAGGACCGTCAATGGGCACGAGTTTGAATAGAGGCGAAGGAGGAGGGA 106
DB	76	TTCACATCTTCGGCTCCGTCAATGGCGCGAGTTCGACCTGGTGGGAGGCGGAAAGGCGCA 135
QY	107	GGCCATACGAAGGCCACATACCTTAAGGTTAAGGTAACCAAGGGGGGACCTTTTGGCCAT 166
DB	136	ACCGAAGCATGG---AACGCTGGAGACCGTGTGAAATCCACCGGGGGCGGCTTGCCT 192
QY	167	TTGCTTGGGATATTTTGTCAACCACATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACC 226
DB	193	GCTCCCGCTGTGTGATCGGACCCCAACTGGGGTACGGTTTCTACCACTACTGCGCCCTTC 252
QY	227	CTGCCGATACACGACTATAAAGCTGTCATTTCTCTGAAGGATTTAAATGGGAAGGG 286
DB	253	CTGCGCGCGCTTCAACCTTCCAAACCCCATCACGACGGGGTACCAAGTTTCAACGTG 312
QY	287	TCATGAAGTTTGAAGACGGTGGCTGCTACTGTAACCCAGGATTCAGATTTCGAGGATG 346
DB	313	TGTTCAAGTTTGAAGACGGGAGTGCTGAATTCGAATCTCCGCTACACCTACGAGGGCG 372
QY	347	GCTGTGTTTCATCTACAAGGTCAAGTTCAATGGCGTGAACCTTTCTTCGATGGACCTGTA 406
DB	373	GCAAGATCAAGGGGAGTTCCAGCTGATCGGTCAGGTTTCCCTGCGCGGGCGCTCTGA 432
QY	407	TG 408
DB	433	TG 434
RESULT 2		
BI386888		681 bp mRNA linear EST 26-AUG-2003
LOCUS		BFL26 001591 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531) Branchiostoma floridae cDNA clone MPMGP531119122 5', mRNA sequence.
DEFINITION		
ACCESSION		BI386888
VERSION		BI386888.1 GI:30921785
KEYWORDS		EST.
SOURCE		Branchiostoma floridae (Florida lancelet)
ORGANISM		Branchiostoma floridae
REFERENCE		Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
AUTHORS		1 (bases 1 to 681)
TITLE		Panopoulou,G., Hennig,S., Groth,D., Krause,A., Poustka,A.J., Herwig,R., Vingron,M. and Lehrach,H.
		New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
JOURNAL		Genome Res. 13 (6A), 1056-1066 (2003)
MEDLINE		22683279
PUBMED		12799346

Contact: Panopoulou G
laboratory 145, dept. Lehrsch
Max-Planck-Institut fuer Molekulare Genetik
Inhnstr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCGATCTTACCATTTATGCTTCGCGTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACGCAGTCGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGTCGGAATTCGCCGGT-3' pSport3/86
High quality sequence stop: 681.
Location/Qualifiers
1. .681
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531I119122"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E. coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention:
BFL26 or MPMGp531)"
/notes="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI,
EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT
primed and directionally cloned in pSport1 vector using a
NotI (5'-PGACTAGTTCAGTCGAGCGCGCC (T)15-3' and a
SalI 5'-TCGACCCACCGCGTCG-3'adapters (Gibco BRL)."

ORIGIN

Query Match	7.2%;	Score 49;	DB 4;	Length 681;
Best Local Similarity	47.8%;	Pred. No. 0.0023;		
Matches 173;	Conservative	0;	Mismatches 186;	Indels 3;
Gaps	1;			

QY 47 TTCGCATGGAAGGAACGGTCAATGGGCACGAGTTTCAAAATAGAAAGCGAAGGAGGGA 106
|||||
Db 87 TTCACATCTTCGGCTCCGTCAATGGCGCGAGTTTCACCTGTTGGGAGGCGGAAGGCA 146
|||||
QY 107 GGCCATACGAAGCCACAATACGTTAAAGTTAAAGTAACCAAGGGGGACCTTTGCCAT 166
|||||
Db 147 ACCCGAAGCATGG---AACGCTCGAGACCAGTGTGAATCCACCCGGGCGCCCTGCCCT 203
|||||
QY 167 TTCCTTTGGGATATTTGTTCACCACAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACC 226
|||||
Db 204 GCTCCCGCTGTGTATCGGACCAACCTGGGTTACCGCTTCTACCAAGTACCTGCCCTTCC 263
|||||
QY 227 CTGCGCACATACAGACTATAAAAGCTGTCAATTTCTCTGAAGGATTAAATGGGAAGGG 286
|||||
Db 264 CTGGCGGCGCTCACCTTCCAAACCGCATACGAGCGGAGGTACCGTTTCCCGTG 323
|||||
QY 287 TCATGAACCTTTGAAGACGGTGGCGTGTACTGTATACCCAGGATTCAGTTTGCAGGATG 346
|||||
Db 324 TGTTCAGTTTGAAGACGGCGGAGTGCTCAATTTGCAACTTCCGCTACACTACNAGGCG 383
|||||
QY 347 GCTGTTTCTACTACAAGGTCAAGTTCATTTGGCGTGAACCTTTCTTCGATGGACCTGTA 406
|||||
Db 384 GCAAGATCAAAAGGGAGTTCCAGCTGATCGGTCAGGTTTCCTTGC CGCGGTCCCGTGA 443
|||||
QY 407 TG 408
|||

Db 444 TG 445

RESULT 3
BI379061
LOCUS
DEFINITION
BFLG1.000692 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498p1084 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 561)
Panopoulou G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTTACATTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1200 Std Error: 0.00
Seq primer: 5'-CCGTCGGATTCGGGT-3' pSport3/86
High quality sequence stop: 561.
Location/Qualifiers
1. 561
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPGp498p1084"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMPGp498)"
/note="vector: pSport1; Site_1: SalI, KpnI, EcoRI (5'); Site_2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5')-pGACTAGTTCTACATCGAGCGCGCC (T)15-3' and a SalI 5'-TCGACCCACCGCTCG-3' adapters (Gibco BRL)."

ORIGIN
Query Match 7.1%; Score 48.4; DB 4; Length 561;
Best Local Similarity 47.8%; Pred. No. 0.0032;
Matches 173; Conservative 0; Mismatches 186; Indels 3; Gaps 1;
47 TTCCATCGAAGAACGGTCAATGGCAGCGATTGTAATAGAGCGGAAGAGAGGGA 106

Db 97 TTCAATCTTCGGCTCGTCAACGGCGCGAGTTCGACCTGGTGGAGGCGGAAGGGCA 156
QY 107 GGCCATACGAAGGCCAATACCGTAAAGCTTAAGGTAAACCAAGGGGGGACCTTTGGCCAT 166
Db 157 ACCGAACGATGG--AACGCTGGAGACCAAGTGTGAATCCACCCGGGGCGCCCTG 213
QY 167 TTGCTTGGGATATTTGTCAACCAATTCAGTATGGAACGAAGGTATATGTCAAGCACC 226
Db 214 GCTCCCGCTGCTGATCGGACCAACCGTGGGGTACGGCTTCTACCACTACCTGCTTCC 273
QY 227 CTGCCGACATACCAAGACTATAAAAGCTGTCAATTTCTCTGAAGGATTTAAATGGGAAGGG 286
Db 274 CTGGCGGCGCTCACCTTCCAAACCGCCATCAGGACGAGGGTACCAAGTTCAAGTG 333
QY 287 TCATGAACCTTTGAAGACGGTGGCTCGTTACTGTAAACCCAGGATTCAGTTTGCAGGATG 346
Db 334 TGTTCAAGTTTGAAGACGGCGGAGTGTGAATTTGCAATTCCTCCGTACACATACGAGGCG 393
QY 347 GCTGTTTCATCTACAGGTCAAGTTCATTTGCGGTGAACCTTCTTCCGATGGACCTGTTA 406
Db 394 GCAAGATCAAAAGGGGAGTTCAGCTGATAGGGTCAGGTTTCCCTGCGCGCGGCTGTGA 453
QY 407 TG 408
Db 454 TG 455

RESULT 4
BI386955
LOCUS
DEFINITION
BFL26.001658 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMPGp531) Branchiostoma floridae cDNA clone MPMPGp531M2297 5', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

1 (bases 1 to 577)
Panopoulou G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTTACATTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1200 Std Error: 0.00
Seq primer: 5'-CCGTCGGATTCGGGT-3' pSport3/86
High quality sequence stop: 561.
Location/Qualifiers
1. 561
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPGp498p1084"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, XL1 blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMPGp498)"
/note="vector: pSport1; Site_1: SalI, KpnI, EcoRI (5'); Site_2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5')-pGACTAGTTCTACATCGAGCGCGCC (T)15-3' and a SalI 5'-TCGACCCACCGCTCG-3' adapters (Gibco BRL)."

ORIGIN
Query Match 7.1%; Score 48.4; DB 4; Length 561;
Best Local Similarity 47.8%; Pred. No. 0.0032;
Matches 173; Conservative 0; Mismatches 186; Indels 3; Gaps 1;
47 TTCCATCGAAGAACGGTCAATGGCAGCGATTGTAATAGAGCGGAAGAGAGGGA 106

Seq primer: 5'-CCGGTCCGGAATTCCTCCGGT-3' pSport3/86
High quality sequence stop: 577.

FEATURES

source
1. 577
Location/Qualifiers
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531M2297"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531)"
/note="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5')-pGACTAGTTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGGGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 7.1%; Score 48.4; DB 4; Length 577;
Best Local Similarity 47.8%; Pred. No. 0.0033;
Matches 173; Conservative 0; Mismatches 186; Indels 3; Gaps 1;
QY 47 TTCGATCGGAAGGAACGGTCAATGGCAGCAGTTTGAATAGAGGCGAAGGAGGGA 106
DB 86 TTACATCTTCGGCTCGTCAATGGCGGAGTTTGGACCTGGTGGAGGTGGAAAGGCA 145
QY 107 GGCCATACGAAGGCCACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCAT 166
DB 146 ACCCGAAGCATGG--AAGCGTGAGACCGAGTGTGAATCCACCGCGCGCTGCCCT 202
QY 167 TTGCTTGGGATATTTTGTACCAACAATTCAGTATGAAGCAAGGTATATGTCAAGCAC 226
DB 203 GCTCCCCGCTGTTGATCGGACCAACCTGGGGTACGGCTTCTACAGTACCTGCCCTTCC 262
QY 227 CTGCGGACATACCAAGCTATATAAAGCTGTCATTCTCTGAAGGATTTAAATGGGAAGGG 286
DB 263 CTGGCGAGCTTACCCTTCCAAACCGCATCAGGAGGAGGTACAGTTCAAGCTG 322
QY 287 TCATGAATTTGAAGACGGTGGCGTCTACTGTAAACCCAGGATTCAGTTTGCAGGATG 346
DB 323 TGTTCAGATTTGAAGACGGCGGAGTGTGAATTCGAATTCCTCCGTACACCTACAAGGGCG 382
QY 347 GCTGTTTCATCTACAAGTCAAGTCAATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 406
DB 383 GCAAGATCAAGGGGAGTTCCAGCTGATCGGGTTCAGGTTTCCCTGCGCGGCTCTGTGA 442
QY 407 TG 408
DB 443 TG 444

RESULT 5
BI386952
LOCUS
DEFINITION
BFL26_001655 Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531) Branchiostoma floridae cDNA clone MPMGP531M0997 5', mRNA sequence.
ACCESSION
VERSION
BI386952
KEYWORDS
SOURCE
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
REFERENCE
1 (bases 1 to 459)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J., Herwig, R., Vingron, M. and Lehrach, H.
TITLE
New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes
JOURNAL
Genome Res. 13 (6A), 1056-1066 (2003)

MEDLINE
PUBMED
COMMENT

22683279
12799346
Contact: Panopoulou G
Laboratory 145, Dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestre 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTTACATTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTAGCCAGCTGGGCAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATTCCTCCGGT-3' pSport3/86
High quality sequence stop: 459.

FEATURES

source
1. 459
Location/Qualifiers
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp531M0997"
/tissue_type="whole embryo"
/dev_stage="26 hrs (neurula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 26hr cDNA library (Name convention: BFL26 or MPMGP531)"
/note="Vector: pSport1 (Gibco BRL); Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); OligodT primed and directionally cloned in pSport1 vector using a NotI (5')-pGACTAGTTAGATCGGAGCGGCC (T)15-3' and a SalI 5'-TCGACCCACGGTCCG-3' adapters (Gibco BRL)."

ORIGIN

Query Match 6.8%; Score 45.8; DB 4; Length 459;
Best Local Similarity 47.2%; Pred. No. 0.019;
Matches 171; Conservative 0; Mismatches 188; Indels 3; Gaps 1;
QY 47 TTCGATCGGAAGGAACGGTCAATGGCAGCAGTTTGAATAGAGGCGAAGGAGGGA 106
DB 84 TTACATCTTCGGCTCGTCAATGGCGGAGTTTCAGCTGGTGGAGGCGGAAGGCA 143
QY 107 GGCCATACGAAGGCCACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCAT 166
DB 144 ACCGGAACGATGG--AAGCGTGAGACCGAGTGTGAATCCACCGCGCGCTGCCCT 200
QY 167 TTGCTTGGGATATTTTGTACCAACAATTTCAAGTATGAAGCAAGGTATATGTCAAGCAC 226
DB 201 GCTCCCGCTGCTGATCGGACCCCAACCTGGGGTACGGCTTCTACAGTACCTGCCCTTCC 260
QY 227 CTGCGGACATACCAAGCTATAAAGCTGTCTTCTCTGAAGGATTTAAATGGGAAGGG 286
DB 261 CTGGCGGCGCTCGCCCTTCCAAACAGCCATCAGGACGAGGATACCAGTTCAAGCTG 320
QY 287 TCATGAATTTGAAGACGGTGGCGTCTGTACTCTGAACCCAGGATTCAGTTTGCAGGATG 346
DB 321 TGTTCAGTTTGAAGACCGCGGAGTGTGAATTCGAATTCCTCCGTACACCTACNAGGCG 380
QY 347 GCTGTTTCATCTACAAGTCAAGTTCATTTGGCGGTGACCTTTCTTCCGATGGACCTGTTA 406
DB 381 GTAAGATCAAGGGGAGTTCAGCTGATCGGGTTCAGGTTTTCCTGCGCGGCGCTGTGA 440


```

QY      407 TG 408
Db      441 TG 442

RESULT 6
BI382373
LOCUS
DEFINITION
BFLG2_000550 Amphioxus 5-6 hrs cDNA library (Name: convention: BFLG
or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498C1177 5',
mRNA sequence.
ACCESSION
BI382373
VERSION
BI382373.1 GI:30917444
KEYWORDS
SOURCE
Branchiostoma floridae (Florida lancelet)
ORGANISM
Branchiostoma floridae
REFERENCE
1 (bases 1 to 591)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genomes Res. 13 (6A), 1056-1066 (2003)
JOURNAL
MEDLINE
PUBMED
22683279
12799346
COMMENT
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (http://www.rzpd.de)
PCR Primers
FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCGAGCTGGCGAAGGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5' -CCGTCGCGAATCCCGGT-3' pSport3/86
High quality sequence stop: 591.
Location/Qualifiers
1..591
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPGp498C1177"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_hosts="E.coli, XLI blue"
/convention="BFLG or MPMPGp498"
/notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSport1 vector using a NotI
(5'-pGACTAGTCTAGATCGGAGCGCCGCC (7)15-3' and a SalI 5'-
TCGACCAACACGCTCCG-3' adaptors (Gibco BRL)."
```

ORIGIN

Query Match 6.7% Score 45.2; DB 4; Length 591;
 Best Local Similarity 47.2%; Pred. No. 0.032;
 Matches 171; Conservative 0; Mismatches 188; Indels 3; Gaps 1;

```

QY      47 TTCCGATGGAAGGAAACGGTCAATGGGCAACGAGTTTGAATAGAACGCGAAGGAGAGGGGA 106
|||
Db      96 TTCACATCTTCGGCTCCGTCATGGCGCGGAGTTTCGACCTGGTGGGAGGCGGAAGGGCA 155
|||
QY      107 GGCATACGAAGGCCCAATACCGTAAAGCTTAAGGTAAACCAAGGGGGGACCTTTGCGCAT 166
|||
Db      156 ACCGGAATGATGG---AACGCTGAGACAGAGTGTGAATCCACCCGGGGGGCGGTGCGCCT 212
|||
QY      167 TTGCTTTGGGATATTTTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACC 226
|||
Db      213 GCTCCCGCTGTTTGTATCGGACCCAACTCGGGGTACGGCTTCTACCACTACCTGCCCCCTCC 272
|||
QY      227 CTGCCGACATACCAGACTATATAAAGCTGTCAATTTCTGTAAGGATTTAATATGGAAAGGG 286
|||
Db      273 CTGGCGGCGCTCACCCTTCCAAACCCGCCATCATCGGACGAGGGGTACAGGTTTCACCGTG 332
|||
QY      287 TCATGAACCTTTGAAGACGGTGGCGTCTTACTCTGAACCCAGGATTCAGTTTCAGGATG 346
|||
Db      333 TGTTCAGTTTGAAGACGGAGGATTTGCTGAATTTCCGCTACCACTACCACTACGAGGGCG 392
|||
QY      347 GCTGTTTTCATCTACAAAGGTCAAGTTTCATTTGGCGTGAACTTTCTTCGATGGACCTGTTA 406
|||
Db      393 GCAAGATCAAAGGGGAGTTCCACCTGATCGGTCAGGTTTCCCTGCGCGGGGCTCTGA 452
|||
QY      407 TG 408
|||
Db      453 TG 454
```

RESULT 7

BI382442

LOCUS

DEFINITION

BFLG2_000619 Amphioxus 5-6 hrs cDNA library (Name: convention: BFLG

or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498C08131 5',

mRNA sequence.

ACCESSION

BI382442

VERSION

BI382442.1 GI:30917513

KEYWORDS

SOURCE

Branchiostoma floridae (Florida lancelet)

ORGANISM

Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;

Branchiostoma.

1 (bases 1 to 707)

Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A. J.,

Herwig, R., Vingron, M. and Lehrach, H.

New evidence for genome-wide duplications at the origin of

vertebrates using an amphioxus gene set and completed animal

genomes

Genomes Res. 13 (6A), 1056-1066 (2003)

JOURNAL

MEDLINE

PUBMED

22683279

12799346

COMMENT

Contact: Panopoulou G

Laboratory 145, dept. Lehrach

Max-Planck-Institut fuer Molekulare Genetik

Innestr. 63-73, D-14195 Berlin, Germany

Tel: +49 30 8413 1235

Fax: +49 30 8413 1128

Email: panopoul@molgen.mpg.de

The library was characterised by oligonucleotide fingerprinting

(ONFP) to reduce sequencing redundancy. According to the ONFP

procedure, clones giving the same hybridisation pattern with a

battery of 200 8mer oligonucleotides are grouped into clusters. One

clone per cluster is selected for sequencing. The size of each

cluster is an indicator of the frequency of a transcript in the

analysed library. The cluster size as well the coordinates of the

rest of the clones assigned to the same fingerprint cluster as the

clone from which the above EST is generated is available at the

amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>

Clones and filters are distributed via the Resource Center/Primary

Database of the German Genome Project (<http://www.rzpd.de>)

PCR Primers

FORWARD: 5' CCCAGGCTTTACACTTATGCTTCGGCTCG 3' (M13RSP)

BACKWARD: 5' GCTATTACCGAGCTGGCGAAGGGGGATGTG 3' (M13FSP)

Insert Length: 1 Std Error: 0.00

Seq primer: 5' -CCGTCGCGAATCCCGGT-3' pSport3/86

High quality sequence stop: 591.

Location/Qualifiers

1..591

/organism="Branchiostoma floridae"

/mol_type="mRNA"

/db_xref="taxon:7739"

/clone="MPMPGp498C1177"

/tissue_type="whole embryo"

/dev_stage="5-6 hrs (gastrula stage)"

/lab_hosts="E.coli, XLI blue"

/convention="BFLG or MPMPGp498"

/notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');"

Site 2: NotI, BamHI, HindIII (3'); OligodT primed and"

directionally cloned in pSport1 vector using a NotI"

(5'-pGACTAGTCTAGATCGGAGCGCCGCC (7)15-3' and a SalI 5'-"

TCGACCAACACGCTCCG-3' adaptors (Gibco BRL)."

BACKWARD: 5' GCTATTACCGCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGTCGCGAATTCCTCGGT-3' pSport3/86
High quality sequence stop: 707.

FEATURES

source
1. .707
Location/Qualifiers
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C08131"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (Gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
/notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligodt primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTTCTAGATCGGAGCGCGCCC (T)15-3' and a SalI 5'-TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."
Query Match 6.6%; Score 44.8; DB 4; Length 707;
Best Local Similarity 46.4%; Pred. No. 0.045;
Matches 182; Conservative 0; Mismatches 207; Indels 3; Gaps 1;
Qy 47 TTGCGATGGAAGGAACGGTCAATGGCGCAGGTTTGAATAGAGGCGAAGGAGGAGGGA 106
Db 141 TGCACATCTATGGCTCATCAACGGTATGGAGTTTGACATGTGGCGGTGGCGATGGCA 200
Qy 107 GGCCATACGAAGGCCACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGGCAT 166
Db 201 ACCAAGAGGCGATCGCTGAGCGTAAACGTGAAGTCTACCAAGGTGCC--TGGCG 257
Qy 167 TTGCTTGGGATATTTTGTACCAAAATTCAGTATGGAAGCAAGGTATATGTCAAGCACC 226
Db 258 TTTCTCTCTGCTGTGGTGGCGCCGATCTAGGGTACGGCCACTACCACTACCTCCCTCC 317
Qy 227 CTGCGGACATACCAAGCTATATAAAGCTGTCATTTCTGGAAGGATTTAAATGGGAAGG 286
Db 318 CTGATGTCGTCGCTTTCCAGGCATCCGTGAACAACGGCGGTATCAATGATCGCT 377
Qy 287 TCATGAATTTGAAGACGGTGGCTGCTTACTGTGAACCCAGGATTCAGTTTGCAGGATG 346
Db 378 CTTTCAACTTTGAGGACGGTGGCTGTGCTGCTGCCACCTACAACCTACTCTACAGCGCG 437
Qy 347 GCTGTTTCATCTACAAGTCAAGTTCATTTGCGTGAACCTTCTTCCGATGGACCTGTTA 406
Db 438 GCAAGATCCAGGGAGAGTTTCATCTGTGGGACAGCTGTTTCCCGACGATAGTCCGCTGA 497
Qy 407 TGCAAAAGAGACAATGGCTGGGAAGCCAGC 438
Db 498 TGACCAAGCGCTGACCGGTCTGGACAGGAGC 529

RESULT 8

BI376909 537 bp mRNA linear EST 26-AUG-2003
LOCUS BFLG_000712 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498) Branchiostoma floridae cDNA clone MPMGP498C24112 5', mRNA sequence.

ACCESSION

BI376909

VERSION

BI376909.1 GI:30911858

KEYWORDS

EST.

SOURCE

Branchiostoma floridae (Florida lancelet)

ORGANISM

Branchiostoma floridae

REFERENCE

1 (bases 1 to 537)

AUTHORS

Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J., Herwig, R., Vingron, M. and Lehrach, H.

TITLE

New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal

JOURNAL
MEDLINE
PUBMED
COMMENT

Genomes

Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoul@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones giving the same hybridisation pattern with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per cluster is selected for sequencing. The size of each cluster is an indicator of the frequency of a transcript in the analysed library. The cluster size as well the coordinates of the rest of the clones assigned to the same fingerprint cluster as the clone from which the above EST is generated is available at the amphioxus project site at: <http://www.molgen.mpg.de/amphioxus/>
Clones and filters are distributed via the Resource Center/Primary Database of the German Genome Project (<http://www.rzpd.de>)
PCR Primers
FORWARD: 5' CCCAGGCTTACATTTATGTTCCGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACCGCAGCTGGCGAAGGGGATGTG 3' (M13FSP)
Insert Length: 1200 Std Error: 0.00
Seq primer: 5'-CCGTCGCGAATTCCTCGGT-3' pSport3/86
High quality sequence stop: 537.

FEATURES

source
1. .537
Location/Qualifiers
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMGp498C24112"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (Gastrula stage)"
/lab_host="E.coli, XLI blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name convention: BFLG or MPMGP498)"
/notes="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5'); Site 2: NotI, BamHI, HindIII (3'); Oligodt primed and directionally cloned in pSport1 vector using a NotI (5'-pGACTAGTTCTAGATCGGAGCGCGCCC (T)15-3' and a SalI 5'-TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."
Query Match 6.4%; Score 43.6; DB 4; Length 537;
Best Local Similarity 47.0%; Pred. No. 0.095;
Matches 170; Conservative 0; Mismatches 189; Indels 3; Gaps 1;
Qy 47 TTGCGATGGAAGGAACGGTCAATGGCGCAGGTTTGAATAGAGGCGAAGGAGGGA 106
Db 79 TGCACATCTTGGCTCCGTCATGGTGGGAGTTCGACCTGTGGGAGGAGGAACGGCA 138
Qy 107 GGCCATACGAAGGCCACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGGCAT 166
Db 139 ACCGGAACGATGG--AACGCTGGAGCAGTGTGAATCCACCGGGCGCTGTCT 195
Qy 167 TTGCTTGGGATATTTTGTACCAACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACC 226
Db 196 GCTTCCCGCTGCTGATCGGACCAACCTGGGGTACGGCTTCTACCACTACCTCCCTCC 255
Qy 227 CTGCGGACATACCAAGCTATATAAAGCTGTCATTTCTGGAAGGATTTAAATGGGAAGG 286
Db 256 CTGGCGGCGCTCGCCCTTCCAAACCGGCATCATTCAGCGAGGTATACAGGTTACCGTG 315
Qy 287 TCATGAATTTGAAGACGGTGGCGTGGTACTGTCTGTAAACCCAGGATTCAGTTTGCAGGATG 346
Db 316 TGTTCAGTTTGAAGACGGAGGAATGCTGAATTTGCAATTTCCCGCTACACCTACAGGGCG 375
Qy 347 GCTGTTTCATCTACAAGGTCAGTTCATTTGCGGTGAACCTTCTTCCGATGGACCTGTTA 406

ORIGIN

Query Match 6.4%; Score 43.6; DB 4; Length 537;
Best Local Similarity 47.0%; Pred. No. 0.095;
Matches 170; Conservative 0; Mismatches 189; Indels 3; Gaps 1;
Qy 47 TTGCGATGGAAGGAACGGTCAATGGCGCAGGTTTGAATAGAGGCGAAGGAGGGA 106
Db 79 TGCACATCTTGGCTCCGTCATGGTGGGAGTTCGACCTGTGGGAGGAGGAACGGCA 138
Qy 107 GGCCATACGAAGGCCACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGGCAT 166
Db 139 ACCGGAACGATGG--AACGCTGGAGCAGTGTGAATCCACCGGGCGCTGTCT 195
Qy 167 TTGCTTGGGATATTTTGTACCAACAATTCAGTATGGAAGCAAGGTATATGTCAAGCACC 226
Db 196 GCTTCCCGCTGCTGATCGGACCAACCTGGGGTACGGCTTCTACCACTACCTCCCTCC 255
Qy 227 CTGCGGACATACCAAGCTATATAAAGCTGTCATTTCTGGAAGGATTTAAATGGGAAGG 286
Db 256 CTGGCGGCGCTCGCCCTTCCAAACCGGCATCATTCAGCGAGGTATACAGGTTACCGTG 315
Qy 287 TCATGAATTTGAAGACGGTGGCGTGGTACTGTCTGTAAACCCAGGATTCAGTTTGCAGGATG 346
Db 316 TGTTCAGTTTGAAGACGGAGGAATGCTGAATTTGCAATTTCCCGCTACACCTACAGGGCG 375
Qy 347 GCTGTTTCATCTACAAGGTCAGTTCATTTGCGGTGAACCTTCTTCCGATGGACCTGTTA 406

```
Db 376 GCAGATCAAGGGAGTTCACCTGCTCGGGTCAGGTTCCCTCCGCGGGCGCTGTGA 435
Qy 407 TG 408
Db 436 TG 437

RESULT 9
BI384961
LOCUS
DEFINITION
BFLG2.003138 Amphioxus 5-6 hrs cDNA library (Name convention: BFLG
or MPMPGp498) Branchiostoma floridae cDNA clone MPMPGp498E05107 5',
mRNA sequence.
BI384961
BI384961.1 GI:30919957
EST.
Branchiostoma floridae (Florida lancelet)
Branchiostoma floridae
Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
Branchiostoma.
1 (bases 1 to 629)
Panopoulou, G., Hennig, S., Groth, D., Krause, A., Poustka, A.J.,
Herwig, R., Vingron, M. and Lehrach, H.
New evidence for genome-wide duplications at the origin of
vertebrates using an amphioxus gene set and completed animal
genomes
Genome Res. 13 (6A), 1056-1066 (2003)
22683279
12799346
Contact: Panopoulou G
Laboratory 145, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Innestr. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1235
Fax: +49 30 8413 1128
Email: panopoulou@molgen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones giving the same hybridisation pattern with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per cluster is selected for sequencing. The size of each
cluster is an indicator of the frequency of a transcript in the
analysed library. The cluster size as well the coordinates of the
rest of the clones assigned to the same fingerprint cluster as the
clone from which the above EST is generated is available at the
amphioxus project site at: http://www.molgen.mpg.de/amphioxus/
Clones and filters are distributed via the Resource Center/Primary
Database of the German Genome Project (http://www.rzpd.de)
PCR Primers
FORWARD: 5' CCCAGGCTTACACTTATGCTTCGGGCTCG 3' (M13RSP)
BACKWARD: 5' GCTATTACGCAGCTGGCGAAGGGGGATGTG 3' (M13FSP)
Insert Length: 1 Std Error: 0.00
Seq primer: 5'-CCGGTCCGGAATTCGGGT-3' pSport3/86
High quality sequence stop: 629.
FEATURES
source
1. .629
/organism="Branchiostoma floridae"
/mol_type="mRNA"
/db_xref="taxon:7739"
/clone="MPMPGp498E05107"
/tissue_type="whole embryo"
/dev_stage="5-6 hrs (gastrula stage)"
/lab_host="E.coli, Xli blue"
/clone_lib="Amphioxus 5-6 hrs cDNA library (Name
convention: BFLG or MPMPGp498)"
/note="Vector: pSport1; Site 1: SalI, KpnI, EcoRI (5');
Site 2: NotI, BamHI, HindIII (3'); OligodT primed and
directionally cloned in pSport1 vector using a NotI
(5')-pGATAGTTAGTCCGAGCGCGGCC (7)15-3' and a SalI 5'-
TCGACCCACGGTCCG-3' adapters (Gibco BRL)."

Query Match 6.4%; Score 43.2; DB 4; Length 629;

ORIGIN
```

```
Best Local Similarity 46.2%; Pred. No. 0.13;
Matches 181; Conservative 0; Mismatches 208; Indels 3; Gaps 1;
Qy 47 TTGCGATGGAAGGAACGGTCAATGGGCACAGTTTGAATAGAAAGCGAAGGAGGGA 106
Db 112 TGCACATCTATGCTCCATCAACGGTATGGAGTTGACATGGTGGCGGTGGAGTGCA 171
Qy 107 GGCATACGAAGGCCACATACCGTAAAGCTTAAGGTAAACCAAGGGGGGACCTTTGCCAT 166
Db 172 ACCCAAGGACCGATCGCTGAGCGTAAACATGAAGTCTACCAAAGGTGCCC---TGGCG 228
Qy 167 TTGCTTGGGATATTTTGTACACCAATTTTCAGTATGGAACCAAGGTATATGTCAAGCAC 226
Db 229 TTTCTCTCTGCTGGTGGGCCCGCATCTGGGGTACGCCACTATCAGTACTACCCCTTCC 288
Qy 227 CTGCCGACATACCGACTATATAAAGCTGTCTTCTCTGAAGGATTTAAATGGGAAGGG 286
Db 289 CTGACGGTCCGTCGCTTCCAGCGAGCTGTGAACACGGCGGTATCAATGATCGCT 348
Qy 287 TCATGAACCTTTGAAGACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTCAGATG 346
Db 349 CTTTCAACTTCGAGGACGGTGGCGTGTGCTGACTGCCACCTACAACTACTCTACAGCGCG 408
Qy 347 GCTGTTTCATCTACAAGTCAAGTTCATTTGCGCGTGAACCTTTCTCCGATGGACCTCTTA 406
Db 409 GCAAGATCCATGGAGAGTTTCATCTGCTGGTGGCGAGCTGTTTCCCGCAGCATGTCGGTGA 468
Qy 407 TGCAAAAGAGACAAATGGGCTGGGAAGCCAGC 438
Db 469 TGACCAACGGCGCTGACCGTTTGGACAGGAGC 500

RESULT 10
LOCUS
DEFINITION
CL065858 930 bp DNA linear GSS 31-DEC-2003
CH216-105J9_RM1.1 CH216 Xenopus tropicalis genomic clone
CH216-105J9, genomic survey sequence.
CL065858
CL065858.1 GI:40521771
GSS.
Xenopus tropicalis (western clawed frog)
Xenopus tropicalis
Xenopus tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus; Silurana.
1 (bases 1 to 930)
Kremitzki, C., Carter, J., McPherson, J., Warren, W., Graves, T.,
Mardis, E. and Wilson, R.
A physical map of the xenopus tropicalis genome
Unpublished (2003)
Contact: Richard K Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Insert Length: 175000 Std Error: 0.00
Seq primer: RM1 TAGACTCACTATAGGAGA
Class: BAC ends
High quality sequence start: 3
High quality sequence stop: 510.
FEATURES
source
1. .930
/organism="Xenopus tropicalis"
/mol_type="genomic DNA"
/strain="Nigerian frog"
/db_xref="taxon:8364"
/clone="CH216-105J9"
/sex="male"
/clone_line="Stock 248 F7A2, inbred N7"
/clone_lib="CH216"
/note="Vector: pTARBAC2.1; CHORI-216 Xenopus tropicalis
BAC library"

ORIGIN
```


/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
 /note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
 Library was constructed by Debopam Chakrabarti.
 Total RNA samples were isolated from mixed stage
 saponin(0.1%) - lysed P. falciparum 3D7 infected
 erythrocytes by the acidic guanidinium-phenol chloroform
 method. The poly A+ RNA was isolated by the polyAT-Tract
 mRNA isolation system (Promega, WI) using streptavidin
 Magnosphere particles. Directional cDNA libraries were
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap
 cDNA synthesis kit (Stratagene, CA). The average size of the
 cDNA inserts in the library was between 1.0 and 1.5kb.
 Clones were mass excised using the ExAssist helper phage
 (Stratagene), the phagemids were precipitated with PEG
 8000 and extracted with phenol/chloroform. Phagemid DNA
 was electroporated into DH10B cells."

ORIGIN

Query Match 5.8%; Score 39.2; DB 4; Length 554;
 Best Local Similarity 55.6%; Pred. No. 2.2;
 Matches 74; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 174 GGATATTTGTCACCAATTTTCAGTATGGAGCAAGGTATATGTCAGGACCCCTGCCGA 233
 |||||
 Db 326 GGATATGNTCAATCCAAATTTGAAGAAATAGGCAAGGAATTTGTTAATCACTACTTTCA 385
 |||||
 QY 234 CATACCAAGCATATAAAGCTGTCTTTCTGGAAGGATTTAAATGGGAAGGGTCATGAA 293
 |||||
 Db 386 GTTATTTAACAGTGGAGGAATGATTTGGCTGCATTATATAAGATATAGTAGTAGAG 445
 |||||
 QY 294 CTTTGAAGACGGT 306
 |||||
 Db 446 TTTTGAAGATGAT 458

RESULT 13

AV719715 592 bp mRNA linear EST 16-OCT-2000
 LOCUS
 DEFINITION AV719715 GLC Homo sapiens cDNA clone GLCGE02 5', mRNA sequence.

ACCESSION AV719715

VERSION AV719715.1 GI:10816867

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 592)
 Qian, B., Wu, T., Huang, Q., Huang, C., Kang, B., Gao, X., Xu, Z.,
 Xiao, H., Xu, X., Li, N., Peng, Y., Liu, F., Qu, J., Song, H., Cheng, Z.,
 Zeng, L., Xu, S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M.,
 Lu, G., Yang, Y., Gu, Y., Chen, Z. and Han, Z.

Homo sapiens cDNA GLC clones
 Unpublished (2000)

TITLE
 JOURNAL
 COMMENT Contact: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China

Tel: 86-21-50801919 (ex. 45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn

This clone is available at CHGC in Shanghai.

FEATURES

source

1. 592
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="GLCGE02"
 /tissue type="corresponding non cancerous liver tissue"
 /dev stage="Adult"
 /lab_host="SOLR"
 /clone_lib="GLC"
 /note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:

XhoI"

ORIGIN
 Query Match 5.8%; Score 39; DB 1; Length 592;
 Best Local Similarity 48.4%; Pred. No. 2.5;
 Matches 108; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 327 GGATTCAGTTTCAGAGATGGCTGTTTCATCTCAAGGTCAGATTTCATGGCGTGAACCTT 386
 |||||
 Db 146 GGATTTTAAATTTTGGAGAGATGAAGGTTTGCCATGGTGCCCAAGCTTCTCTTGAACCTT 205
 |||||
 QY 387 TCCTTCGATGGACCTGTTTATGCAAAAGAGACAATGGGCTGGAGCCAGCAGCTGAGCG 446
 |||||
 Db 206 TTGGACTCAGCAATTTCTTCTGGCTTAGACTTCCAAGGGCTGGAAATTCACGGCTGAGCT 265
 |||||
 QY 447 TTTGTATCTCTGGTATGGCGTGTGAAAGGAGAGATTCATTAAGGCTCTTGAAGCTGAAAGA 506
 |||||
 Db 266 ACTGTGCCCCCCCCCTACTTTTAAAAATAACAAATGTGATGGTATTTTCCAGACACG 325

RESULT 14

BZ131346

LOCUS

DEFINITION BZ131346 808 bp DNA linear GSS 11-OCT-2002

CH230-345J23.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone

BZ131346 genomic survey sequence.

ACCESSION BZ131346

VERSION BZ131346.1 GI:23772293

KEYWORDS GSS.

SOURCE Rattus norvegicus (Norway rat)

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 808)

AUTHORS Zhao, S., Shetty, J., Shatsman, S., Teegaye, G., Geer, K.,

Shvartsbeyn, A., Gobregeorgis, E., Overton, L., Russell, D.,

Riggs, F., de Jong, P. and Fraser, C.M.

Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)

Other GSSs: CH230-345J23.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pdejong@mail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or ering information.htm). BAC end

plate: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html

Seq primer: SP6

Class: BAC ends.

Location/Qualifiers

1. 808

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="BN/SENHsd/MCW"

/db_xref="taxon:10116"

/clone="CH230-345J23"

/sex="Female"

/cell type="Brain"

/clone_lib="CHORI-230 Segment 2"

/notes="Vector: PTABAC1.3; Site_1: MboI; Site_2: MboI;

CHORI-230 Rat (BN/SENHsd/MCW) BAC library produced by

Pieter de Jong"

ORIGIN

Query Match 5.8%; Score 39; DB 8; Length 808;
Best Local Similarity 56.7%; Pred. No. 2.8;
Matches 72; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 434 CCAGCACTGACGCTTGTATCTCGTGATGCGGTGTTGAAGAGGAGATTCAATGAGCTC 493
|||||
Db 256 CCAGCACTGCTGCTTGCATGACCAAGCTTCCTGCCATGATGATAATAGACTAAGCCTC 315
|||||

Qy 494 TGAAGCTGAAGACGGTGGTCAATACCTAGTTGAATTCAMAAGTATTACATGCGCAAGA 553
|||||
Db 316 TGAATTTGTAAGCCAGCCCAATTAATGTTTGTCTTTATAAGAGTTTCCATGTTAAGG 375
|||||

Qy 554 AGCCTGT 560
|||||
Db 376 TGCCTCT 382
|||||

ORIGIN

Query Match 5.7%; Score 38.6; DB 5; Length 571;
Best Local Similarity 55.6%; Pred. No. 3.3;
Matches 74; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 174 GGATATTTTGTACCAACAATTTTCAGTATGGAAGCAAGGTATATGTCAGCACCTGCCGA 233
|||||
Db 211 GGATATGCTCAATCCACAATTTGAAGAAATAGCAAGGAATTTGTTAATCACTACTTTCA 270
|||||

Qy 234 CATACCAGACTATAAAAAGCTGTCATTTCTCTGAAGGATTTAAATGGAAAGGGTCATGAA 293
|||||
Db 271 GTTATTTACAGTGAAGGAAATGAATGGCTGCATTATATAAGATATAAGTATGATGAG 330
|||||

Qy 294 CTTTGAAGACGGT 306
|||||
Db 331 TTTTGAAGATGAT 343
|||||

Search completed: July 1, 2005, 17:24:20
Job time : 3048 secs

RESULT 15
B0451564
LOCUS
DEFINITION Pf8Stoab06b02.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum 3D7 cDNA 5' similar to SW:NTF2_NEUCR P87102 NUCLEAR TRANSPORT FACTOR 2 ; mRNA sequence.

ACCESSION B0451564
VERSION B0451564.1 GI:21254676
KEYWORDS EST.

SOURCE
ORGANISM

REFERENCE
AUTHORS

1 (bases 1 to 571)
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D., Marra, K., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I., Teagareishvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A., Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D., Waterston, R., Wilson, R. and Sibley, D.

WashU Plasmodium EST Project
Unpublished (2001)

Contact: L. David Sibley

WashU Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library was constructed by Debopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -40UP from Gibco

High quality sequence stop: 420.

Location/Qualifiers

FEATURES
source

1..571
/organism="Plasmodium falciparum 3D7"
/mol_type="mRNA"
/db_xref="taxon:36329"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
/note="vector: pBluescript SK plus; Site 1: EcoRI; Site 2: XhoI; Library was constructed by Debopam Chakrabarti. Total RNA samples were isolated from mixed stage saponin(0.1%) -lysed P. falciparum 3D7 infected erythrocytes by the acidic guanidinium-phenol chloroform method. The poly A+ RNA was isolated by the polyAT-Tract mRNA isolation system (Promega, WI) using streptavidin Magnosphere particles. Directional cDNA libraries were constructed by oligo d(T) priming of poly(A)+ RNA (5mg) into EcoRI and XhoI sites of 1 ZapII vector using the Zap cDNA synthesis kit (Stratagene, CA). The average size of the cDNA inserts in the library was between 1.0 and 1.5kb. Clones were mass excised using the ExAssist helper plasmid

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 12:17:58 ; Search time 1029 Seconds
(without alignments)
3900.471 Million cell updates/sec

Title: US-10-006-922a-11
Perfect score: 678
Sequence: 1 atgaggtcttccaagaatgt.....gccaccatctgttcctttaa 678

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 16Dec04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678	100.0	678	4	Aah47654 Discosoma
2	678	100.0	678	4	Aad13053 Discosoma
3	678	100.0	678	4	Aad11142 Discosoma
4	678	100.0	678	6	Aba95921 Yeast Opt
5	678	100.0	678	10	Adc24127 Discosoma
6	678	100.0	678	11	Adl46204 Discosoma
7	678	100.0	678	11	Adn33978 Wild-type
8	678	100.0	678	12	Adi36420 Discosoma
9	678	100.0	678	12	Adm97768 D sp red
10	678	100.0	859	6	Aal47952 Discosoma
11	678	100.0	859	8	Aad53432 Discosoma
12	678	100.0	859	10	Aad61969 Discosoma
13	677	99.9	678	10	Adf70404 Discosoma
14	677	99.9	3311	9	Aca62995 Plasmid D
15	665	98.1	666	6	AbL61142 Red Fluor
16	661	97.5	711	12	Ado78067 Corallimo
17	661	97.5	711	12	Ado78069 Corallimo
18	658.8	97.2	898	4	Aad03615 Discosoma
19	657.2	96.9	898	3	Aad48742 Discosoma
20	642.8	94.8	678	12	Adm13536 DNA encod

21	566.4	83.5	876	3	AAA28958
22	566.4	83.5	876	4	AAD03618
23	566.4	83.5	876	10	ACC44649
24	450.6	66.5	723	8	ABZ22476
25	449.8	66.3	681	4	AAH47656
26	449.2	66.3	678	6	AAD46278
27	449.2	66.3	678	6	AAD28207
28	449.2	66.3	695	3	AAA48743
29	447.6	66.0	678	6	ABA95922
30	447.2	66.0	681	10	ADC24134
31	447.2	66.0	681	11	ADL46205
32	447.2	66.0	681	11	ADL46225
33	447.2	66.0	2721	13	ADS75466
34	447.2	66.0	2772	13	ADS75468
35	447.2	66.0	4692	6	AAL47954
36	447.2	66.0	4692	10	ACC44640
37	447.2	66.0	7910	4	AAD09979
38	447.2	66.0	9320	6	ABS56664
39	447.2	66.0	9658	12	ADI04104
40	447.2	66.0	9678	12	ADI04103
41	447.2	66.0	10263	12	ADI04102
42	446	65.8	678	6	AAD28208
43	445.8	65.8	1638	10	ADL18131
44	445.8	65.8	1647	10	ADL18155
45	445.8	65.8	6893	10	ADE24111

ALIGNMENTS

RESULT 1
AAH47654
ID AAH47654 standard; cDNA; 678 BP.
XX
AC AAH47654;
XX
DT 30-NOV-2001 (first entry)
XX
DE Discosoma sp. red drFP583 protein coding sequence.
XX
KW Fluorescent protein; Anthozoan; fluorescence; marker; FRET; drFP583; ss.
OS Discosoma sp.
XX
PF WO200162919-A1.
PN XX
PR 30-AUG-2001.
XX
PF 13-FEB-2001; 2001WO-US004625.
XX
PR 23-FEB-2000; 2000US-0184732P.
XX
PA (AURO-) AURORA BIOSCIENCES CORP.
XX
PI Nelson D, Zamaira E, Tsien R;
XX
WPI; 2001-557704/62.
XX
PT Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise functional red fluorescent proteins, and the encoding nucleic acids, with key mutations for improving the proteins function.
XX
PS Disclosure; Page 83; 90pp; English.
XX
CC The invention provides a nucleic acid encoding functional red fluorescent CC protein (II) that differs from the sequence of an Anthozoan red CC fluorescent protein by at least one amino acid substitution, and with CC different fluorescent properties. The red fluorescent protein of (II) are CC invention can be expressed by standard recombinant methodology. (II) are CC used a fluorescent markers and FRET partners. It is used for identifying CC protein-protein interactions. (II) is also suitable for multiplexed CC fluorescent analysis and FRET-based applications using existing Aequorea CC fluorescent proteins. (II) has improved brightness, reduced spectral

```
CC cross talk, and is rapidly and efficiently expressed in mammalian cells.
CC The key mutations in the encoding nucleic acids provide improved folding,
CC brightness, and create (11) with sharper, more defined excitation and
CC emission peaks when expressed in mammalian cells. The present sequence
CC represents the coding sequence of a Discosoma sp. drPP583 protein, an
CC anthozoan fluorescent protein
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 4; Length 678;
Best Local Similarity 100.0%; Pred. No. 9.8e-208;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGTCCTCCAAAGAAATGTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGGA 60
DB |||||||
QY 1 ATGAGGTCCTCCAAAGAAATGTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGGA 60
DB |||||||
QY 61 ACGGTCAATGGGCACGAGTTTGAATAGAAAGGCGAAGAGAGGGAGGCCATACGAAGGC 120
DB |||||||
QY 61 ACGGTCAATGGGCACGAGTTTGAATAGAAAGGCGAAGAGAGGGAGGCCATACGAAGGC 120
DB |||||||
QY 121 CACAATACCGTAAAGCTTAAAGGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB |||||||
QY 121 CACAATACCGTAAAGCTTAAAGGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB |||||||
QY 181 TTGTCAACCAAAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTCGCGACATACCA 240
DB |||||||
QY 181 TTGTCAACCAAAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTCGCGACATACCA 240
DB |||||||
QY 241 GACTATAAAAAGCTGTCTTCTGGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB |||||||
QY 241 GACTATAAAAAGCTGTCTTCTGGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB |||||||
QY 301 GACGGTGGCGTCGTACTGCTAAACCCAGGATTCAGATTTCAGATTTCAGATTTCAGATTTC 360
DB |||||||
QY 301 GACGGTGGCGTCGTACTGCTAAACCCAGGATTCAGATTTCAGATTTCAGATTTCAGATTTC 360
DB |||||||
QY 361 AAGGTCAAGTTCATTGCGCGTGAACCTTTCCTTCCGATGGACCTGTATGCAAAAGAGACA 420
DB |||||||
QY 361 AAGGTCAAGTTCATTGCGCGTGAACCTTTCCTTCCGATGGACCTGTATGCAAAAGAGACA 420
DB |||||||
QY 421 ATGGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
DB |||||||
QY 421 ATGGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
DB |||||||
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTCGAATTCBAAGTATT 540
DB |||||||
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTCGAATTCBAAGTATT 540
DB |||||||
QY 541 TACATGGCAAGAGAGCCCTGTGCAAGTACCAGGGTACTTACTATGTTGACTCCAAACTGGAT 600
DB |||||||
QY 541 TACATGGCAAGAGAGCCCTGTGCAAGTACCAGGGTACTTACTATGTTGACTCCAAACTGGAT 600
DB |||||||
QY 601 ATAAAGGCCACACGAAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
DB |||||||
QY 601 ATAAAGGCCACACGAAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
DB |||||||
QY 661 CACCATCTGTTCTTTAA 678
DB |||||||
QY 661 CACCATCTGTTCTTTAA 678
DB |||||||

RESULT 2
AAD13053
ID AAD13053 standard; cDNA; 678 BP.
XX
AC AAD13053;
XX
DT 16-OCT-2001 (first entry)
XX
DE Discosoma sp. "red" anthozoa fluorescent protein, drPP583 cDNA.
XX
KW Protein destabilisation; linker moiety; reporter moiety; disease model;
```

```
KW linear multimerised domain; -NH-ubiquitin protein endoprotease;
KW transgenic animal; transgenic plant; disease resistance;
KW anthozoa fluorescent protein; natural fluorescent protein; ss.
XX
OS Discosoma sp.
XX WO200157242-A2.
XX
PN 09-AUG-2001.
XX
PF 02-FEB-2001; 2001WO-US003791.
XX
PR 04-FEB-2000; 2000US-00498098.
XX
PA (AURO-) AURORA BIOSCIENCES CORP.
XX
PI Stack JH, Whitney M, Cubitt AB, Pollok BA;
XX WPI; 2001-488890/53.
XX
PT Destabilizing proteins in living cells, by coupling a target protein to
PT linear multimerized destabilization domain non-cleavable by -NH-ubiquitin
PT protein endoproteases, comprising two copies of the domain.
XX
PS Disclosure; Page 110; 17ipp; English.
XX
CC The present invention relates to a method for destabilising a target
CC protein in a cell. The method comprises a linker moiety which operatively
CC couples a target protein (a reporter moiety) to a linear multimerised
CC destabilising domain, which is non-cleavable by a -NH-ubiquitin protein
CC endoprotease. The method is useful for detecting an activity such as
CC protease, protein kinase or phosphoprotein phosphatase activity and is
CC also useful for identifying modulators of these activities. The method is
CC also useful for developing novel assays for a wide range of post-
CC translational activities, such as proteolysis, phosphorylation,
CC dephosphorylation, glycosylation, methylation, sulfation, prenylation,
CC disulfide bond formation and ADP-ribosylation within cells. The
CC recombinant DNA molecule of the invention is useful for creating
CC transgenic animals useful as disease models and transgenic plants with
CC improved disease resistance or other favourable traits. The present
CC sequence is Discosoma sp. "red" anthozoa fluorescent protein, drPP483
CC cDNA which is a natural fluorescent protein used as a reporter moiety in
CC the exemplification of the invention
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 4; Length 678;
Best Local Similarity 100.0%; Pred. No. 9.8e-208;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGTCCTCCAAAGAAATGTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGGA 60
DB |||||||
QY 1 ATGAGGTCCTCCAAAGAAATGTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGGA 60
DB |||||||
QY 61 ACGGTCAATGGGCACGAGTTTGAATAGAAAGGCGAAGAGAGGGAGGCCATACGAAGGC 120
DB |||||||
QY 61 ACGGTCAATGGGCACGAGTTTGAATAGAAAGGCGAAGAGAGGGAGGCCATACGAAGGC 120
DB |||||||
QY 121 CACAATACCGTAAAGCTTAAAGGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB |||||||
QY 121 CACAATACCGTAAAGCTTAAAGGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB |||||||
QY 181 TTGTCAACCAAAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTCGCGACATACCA 240
DB |||||||
QY 181 TTGTCAACCAAAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTCGCGACATACCA 240
DB |||||||
QY 241 GACTATAAAAAGCTGTCTTCTGGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB |||||||
QY 241 GACTATAAAAAGCTGTCTTCTGGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB |||||||
QY 301 GACGGTGGCGTCGTACTGCTAAACCCAGGATTCAGATTTCAGATTTCAGATTTCAGATTTC 360
DB |||||||
QY 301 GACGGTGGCGTCGTACTGCTAAACCCAGGATTCAGATTTCAGATTTCAGATTTCAGATTTC 360
DB |||||||
QY 361 AAGGTCAAGTTCATTGCGCGTGAACCTTTCCTTCCGATGGACCTGTATGCAAAAGAGACA 420
DB |||||||
QY 361 AAGGTCAAGTTCATTGCGCGTGAACCTTTCCTTCCGATGGACCTGTATGCAAAAGAGACA 420
DB |||||||
QY 421 ATGGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
DB |||||||
QY 421 ATGGGCTGGGAAGCAGCAGCTAGCGTTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
DB |||||||
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTCGAATTCBAAGTATT 540
DB |||||||
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTCGAATTCBAAGTATT 540
DB |||||||
QY 541 TACATGGCAAGAGAGCCCTGTGCAAGTACCAGGGTACTTACTATGTTGACTCCAAACTGGAT 600
DB |||||||
QY 541 TACATGGCAAGAGAGCCCTGTGCAAGTACCAGGGTACTTACTATGTTGACTCCAAACTGGAT 600
DB |||||||
QY 601 ATAAAGGCCACACGAAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
DB |||||||
QY 601 ATAAAGGCCACACGAAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
DB |||||||
QY 661 CACCATCTGTTCTTTAA 678
DB |||||||
QY 661 CACCATCTGTTCTTTAA 678
DB |||||||

RESULT 2
AAD13053
ID AAD13053 standard; cDNA; 678 BP.
XX
AC AAD13053;
XX
DT 16-OCT-2001 (first entry)
XX
DE Discosoma sp. "red" anthozoa fluorescent protein, drPP583 cDNA.
XX
KW Protein destabilisation; linker moiety; reporter moiety; disease model;
```


QY 361 AAGTCAAGTTCAATGGGCGAAGCTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
DB |||||||
QY 361 AAGGTCAAGTTCAATGGGCGAAGCTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
DB |||||||
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGCGTGTGAAAGGAGAG 480
DB |||||||
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGCGTGTGAAAGGAGAG 480
DB |||||||
QY 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
DB |||||||
QY 541 TACATGGAAGAGAGCTGTGACGATACCGGTTACTATGTTGATCTCAAACTGGAT 600
DB |||||||
QY 541 TACATGGAAGAGAGCTGTGACGATACCGGTTACTATGTTGATCTCAAACTGGAT 600
DB |||||||
QY 601 ATAAAGAGCCCAACGAGACTATACAACTGTTGAGCAGTATGAAGAACCGGGGAGCG 660
DB |||||||
QY 661 CACCATCTGTTCCCTTTAA 678
DB |||||||
QY 661 CACCATCTGTTCCCTTTAA 678
DB |||||||

RESULT 3

AAD11142
ID AAD11142 standard; DNA; 678 BP.
XX
AC AAD11142;
XX
DT 24-SEP-2001 (first entry)
XX
DE Discosoma species Anthozoa fluorescent protein, drFP583 encoding DNA.
XX
KW Transmembrane potential; biological membrane; fluorescent ion; detection;
KW test chemical screening; Anthozoa fluorescent protein; FP;
KW transgenic organism; drFP583 protein; ds.
XX
OS Discosoma sp.
XX
PN WO200142211-A2.
XX
PD 14-JUN-2001.
XX
PF 12-DEC-2000; 2000WO-US033739.
XX
PR 13-DEC-1999; 99US-00459956.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI Tsien RY, Gonzalez JE;
XX
DR WPI; 2001-457276/49.
XX
PT Determining electrical potential across a membrane in biological systems,
PT comprises introducing two reagents, exposing the membrane to light and
PT measuring the energy transfer.
XX
PS Disclosure; Page 150; 154pp; English.
XX
CC The patent discloses optical methods and compositions for determining
CC transmembrane potential across biological membranes in living cells. The
CC method of determining the electrical potential across a membrane in a
CC biological system comprises introduction of two reagents, a first reagent
CC comprising a hydrophobic fluorescent ion capable of redistributing from a
CC first face of the membrane to a second face of the membrane in response
CC to membrane potential change and a second reagent that label the first
CC face or the second face of the membrane, which comprises a chromophore
CC capable of undergoing energy transfer by either donating or accepting
CC excited state energy to the fluorescent ion. The membrane is then exposed
CC to excitation light and the energy transfer between the reagents is

CC measured and related to the membrane potential. The method is useful for
CC detecting changes in membrane potential in subcellular organelle
CC membranes in biological systems. The method is used for screening of test
CC chemicals for activity to modulate the activity of target ion channel.
CC The invention also provides a transgenic organism comprising a first
CC reagent that comprises a charged hydrophobic fluorescent molecule and a
CC second reagent comprising a bioluminescent or naturally fluorescent
CC protein. The present sequence is Discosoma species (red), DNA encoding an
CC Anthozoa fluorescent protein (FP), drFP583
XX

SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 4; Length 678;
Best Local Similarity 100.0%; Pred. No. 9.8e-208;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGGTCTTCCAAAGTATTATCAAGGAGTTTCAAGGTTTAAAGTTTCGCATGGAAGGA 60
DB |||||||
QY 1 ATGAGGTCTTCCAAAGTATTATCAAGGAGTTTCAAGGTTTAAAGTTTCGCATGGAAGGA 60
DB |||||||
QY 61 ACGGTCAATGGGCAAGCGAGTTTGAATAGAAAGGCGAAGAGAGGGAGGCCATACGAAGGC 120
DB |||||||
QY 61 ACGGTCAATGGGCAAGCGAGTTTGAATAGAAAGGCGAAGAGAGGGAGGCCATACGAAGGC 120
DB |||||||
QY 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGGACCTTTGCCATTTGCTGGGATATT 180
DB |||||||
QY 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGGACCTTTGCCATTTGCTGGGATATT 180
DB |||||||
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTCGCGACATACCA 240
DB |||||||
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTCGCGACATACCA 240
DB |||||||
QY 241 GACTATATAAAGCTGTCTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
DB |||||||
QY 301 GACGTGGCGTCTGTTACTGTAAACCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
DB |||||||
QY 301 GACGTGGCGTCTGTTACTGTAAACCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
DB |||||||
QY 361 AAGGTCAAGTTCAATGGGCGTGAAGCTTTCCTTCGATGGACCTGTATGCAAAAGAGACA 420
DB |||||||
QY 361 AAGGTCAAGTTCAATGGGCGTGAAGCTTTCCTTCGATGGACCTGTATGCAAAAGAGACA 420
DB |||||||
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGCGTGTGAAAGGAGAG 480
DB |||||||
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGCGTGTGAAAGGAGAG 480
DB |||||||
QY 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
DB |||||||
QY 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
DB |||||||
QY 541 TACATGGAAGAGAGCTGTGACGATACCGGTTACTATGTTGATCTCAAACTGGAT 600
DB |||||||
QY 541 TACATGGAAGAGAGCTGTGACGATACCGGTTACTATGTTGATCTCAAACTGGAT 600
DB |||||||
QY 601 ATAAAGAGCCCAACGAGACTATACAACTGTTGAGCAGTATGAAGAACCGGGGAGCG 660
DB |||||||
QY 661 CACCATCTGTTCCCTTTAA 678
DB |||||||
QY 661 CACCATCTGTTCCCTTTAA 678
DB |||||||

RESULT 4

ABA95921
ID ABA95921 standard; DNA; 678 BP.
XX
AC ABA95921;
XX
DT 29-MAY-2002 (first entry)
XX

CC in reduced propensity of the DsRed variant to form tetramers. (I) is
CC useful for detecting transcriptional activity by providing a host cells
CC containing a vector which comprises (I) operatively linked to an
CC expression control sequence, and an unit to assay the variant fluorescent
CC protein fluorescence, and assaying fluorescence of the variant
CC fluorescent protein produced by (VII), where variant fluorescent protein
CC fluorescence is indicative of transcriptional activity. A polynucleotide
CC encoding a fusion protein is useful for the analysis of in vivo
CC localisation or trafficking of a polypeptide of interest. A polypeptide
CC marker is useful as markers to identify the location and amount of a
CC target protein produced, where the target protein is fused to the marker,
CC as a complement to or alternative for the green fluorescent protein or
CC its spectral variant, for detecting induction of transcriptions, in
CC applications involving fluorescence energy resonance transfer (FRET),
CC which detects events as the function of the movement of fluorescent
CC donors and acceptors towards or away from each other, for making
CC fluorescent sensors for protein kinase and phosphatase activities or
CC indicators for ions and molecules such as Ca^{2+} , Zn^{2+} , for identifying the
CC presence of a molecule in a sample, for identifying a specific
CC interaction of a first and second molecule, for determining whether a
CC sample contains an enzyme or for determining the pH of the sample. (I) is
CC useful for identifying a region or condition that regulates the activity
CC of an expression control sequence. This sequence encodes Discosoma wild-
CC type red fluorescent protein.

XX SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 10; Length 678;

Best Local Similarity 100.0%; Pred. No. 9.8e-208;

Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATCGAAGGA 60
Db 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATCGAAGGA 60
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGGGGAGGCCATACGAAGGC 120
Qy 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGAGCCCTTGGCCATTTGCTTGGGATATT 180
Db 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGAGCCCTTGGCCATTTGCTTGGGATATT 180
Qy 181 TTGTACACAAATTTCTAGTATGGAAGCAAGGTATATGTCAAGCACCTGCCGACATACCA 240
Db 181 TTGTACACCAATTTCTAGTATGGAAGCAAGGTATATGTCAAGCACCTGCCGACATACCA 240
Qy 241 GACTATAAAGAGCTGTCATTTCTGAGAGTTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATAAAGAGCTGTCATTTCTGAGAGTTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGATTTCAGAGTGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGATTTCAGAGTGGCTGTTTCATCTAC 360
Qy 361 AAGGTCAAGTTTCATTTGGCGTGAACCTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCAAGTTTCATTTGGCGTGAACCTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTGTATGCGCTGTTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTGTATGCGCTGTTGAAAGGAGAG 480
Qy 481 ATTCTAATAGGCTCTGAAGCTGAAGAGCGTGGTCTATTACCTAGTTGTAATCAAAAGTATT 540
Db 481 ATTCTAATAGGCTCTGAAGCTGAAGAGCGTGGTCTATTACCTAGTTGTAATCAAAAGTATT 540
Qy 541 TACATGGCAAGAAGCCCTGTGAGCTACACGGGTACTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAGAAGCCCTGTGAGCTACACGGGTACTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATACAGCCCAACAGAACTATACAACTGTTGAGCGATGTAAAGAACCGAGGGACGC 660
Db 601 ATACAGCCCAACAGAACTATACAACTGTTGAGCGATGTAAAGAACCGAGGGACGC 660

Db 601 ATACAGCCCAACAGAACTATACAACTGTTGAGCGATGTAAAGAACCGAGGGACGC 660
Qy 661 CACCATCTGTTCTTTAA 678
Db 661 CACCATCTGTTCTTTAA 678

RESULT 6

ADL46204

XX ADL46204 standard; DNA; 678 BP.

AC ADL46204;

XX 20-MAY-2004 (first entry)

DE Discosoma red fluorescent protein (DsRed) coding sequence.

XX ds; gene; red fluorescent protein; DsRed; fluorescence; red wavelength;
XX oligomerization; tetramerization; immunoassay; hybridization assay.

XX Discosoma sp.

PH Key Location/Qualifiers

FT CDS 1..678

FT /*tag= a

FT /product= "DsRed protein"

XX WO2003086446-A1.

XX 23-OCT-2003.

XX 09-APR-2003; 2003WO-US010879.

XX 10-APR-2002; 2002US-00121258.

XX 29-JUL-2002; 2002US-00209208.

XX (REGC) UNIV CALIFORNIA.

XX Tsien RY, Campbell RE, Baird GS;

XX WPI; 2003-845265/78.

XX P-PSDB; ADL46203.

XX New monomeric and dimeric Anthozoan fluorescent protein variants with
XX reduced propensity to oligomerize, and encoding polynucleotides, useful
XX in molecular biology, e.g. in immunoassays or in tracking protein
XX movement in cells.

XX Disclosure; SEQ ID NO 2; 166pp; English.

XX The invention relates to a polynucleotide sequence encoding a Discosoma
XX red fluorescent protein (DsRed) variant having a reduced propensity to
XX oligomerize. The protein variant comprises one or more amino acid
XX substitutions at the AB and/or AC interface(s) of the wild-type DsRed
XX sequence, where the substitutions result in reduced propensity of the
XX DsRed variant to form tetramers and where the variant displays detectable
XX fluorescence of at least one red wavelength. The composition and methods
XX are useful in producing red fluorescent proteins having reduced
XX propensity for oligomerization, especially tetramerization. The protein
XX may be used in molecular biology and in other scientific applications,
XX such as in immunoassays or hybridization assays, or in tracking the
XX movement of proteins in cells. This sequence corresponds to the DsRed
XX coding sequence.

XX Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 11; Length 678;

Best Local Similarity 100.0%; Pred. No. 9.8e-208;

Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATCGAAGGA 60

Db 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTTCATGAGTTTAAAGTTTCGATCGAAGGA 60

Qy 61 ACGGTCATAGGCGACGCTTTGAATAAGCGCGAAGGAGAGGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCATAGGCGACGCTTTGAATAAGCGCGAAGGAGAGGGGAGGCCATACGAAGGC 120
Qy 121 CACAATACCGTAAAGCTTAAAGGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAAGGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Qy 181 TTGTCACCAACAATTCAGTATGGAAGCAAGAGTATATGTCAAGCACCTTCCCGCATACCA 240
Db 181 TTGTCACCAACAATTCAGTATGGAAGCAAGAGTATATGTCAAGCACCTTCCCGCATACCA 240
Qy 241 GACTATATAAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATATAAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 301 GACGGTGGCGTGGTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTGGTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGTCAAGTTCATGGCGTGAACCTTCTTCCGATGGACCTGTATGCAAAAGAGACA 420
Db 361 AAGTCAAGTTCATGGCGTGAACCTTCTTCCGATGGACCTGTATGCAAAAGAGACA 420
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 480
Qy 481 ATTCATAGGCTGTGAAGCTGAAGAGACGGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Db 481 ATTCATAGGCTGTGAAGCTGAAGAGACGGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Qy 541 TACATGGCAAGAAGCCCTGTGAGCTACAGGGTACTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAGAAGCCCTGTGAGCTACAGGGTACTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATAAACGCCACAAGAGACTATACATCGTTGAGCAGTATGAAGAACCAGGAGCGC 660
Db 601 ATAAACGCCACAAGAGACTATACATCGTTGAGCAGTATGAAGAACCAGGAGCGC 660
Qy 661 CACCATCTGTTCCCTTTAA 678
Db 661 CACCATCTGTTCCCTTTAA 678

RESULT 7

ID ADN33978
AC ADN33978 standard; DNA; 678 BP.
XX ADN33978;
XX
DT 01-JUL-2004 (first entry)
XX
DE Wild-type DsRED encoding sequence.
XX
KW Cnidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED;
KW ds.
XX
OS Discosoma sp.
XX
PN WO2003054158-A2.
XX
PD 03-JUL-2003.
XX
PF 18-DEC-2002; 2002WO-US040539.
XX
PR 19-DEC-2001; 2001US-0341723P.
XX
PA (UYCH-) UNIV CHICAGO.
XX
FI Bevis B, Glick B;
XX

DR WPI; 2003-569236/53.
DR P-PSDB; ADN33979.
XX
PT Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent
PT mutant of a Cnidarian chromo- or fluorescent protein or its mutant,
PT useful for applications involving chromo- or fluorescent proteins.
XX
PS Claim 8; SEQ ID NO 1; 65pp; English.
XX
CC The present invention relates to nucleic acid that encodes a rapidly
CC maturing chromo or fluorescent mutant of a Cnidarian chromo- or
CC fluorescent protein or its mutant. The protein is useful in applications
CC involving nucleic acid encoding a chromo- or fluorescent protein and is
CC useful for producing a chromo and/or fluorescent protein which involves
CC growing the cell, whereby the protein is expressed, and isolating the
CC protein substantially free of other proteins. The protein is useful in
CC applications involving chromo- or fluorescent protein and is useful as
CC PCR primers, hybridization probes, etc. The expression cassettes are
CC useful for synthesizing related proteins. The chromoproteins are useful
CC as coloring agents which are capable of imparting color or pigment to a
CC particular composition of matter e.g. food compositions, pharmaceuticals,
CC cosmetics, living organisms, e.g., animals and plants. The chromoproteins
CC may also find use as labels in analyte detection assays, e.g. assays for
CC biological analytes of interest and as selectable markers in recombinant
CC DNA applications, e.g. the production of transgenic cells and organisms.
CC The fluorescent proteins find use in a variety of different applications,
CC e.g. in fluorescence resonance energy transfer (FRET) applications, as
CC biosensors in prokaryotic and eukaryotic cells, in applications involving
CC the automated screening of arrays of cells expressing fluorescent
CC reporting groups by using microscopic imaging and electronic analysis, as
CC second messenger detectors, and in fluorescence activated cell sorting
CC applications and as in vivo marker in animals. The fluorescent proteins
CC also find use in protease cleavage assays. The proteins can also be used
CC is assays to determine the phospholipid composition in biological
CC membranes and as a fluorescent timer. The present sequence represents the
CC wild-type DsRED encoding sequence.
XX
SQ Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;
Query Match 100.0%; Score 678; DB 11; Length 678;
Best Local Similarity 100.0%; Pred. No. 9.8e-208;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGAGGTCTTCCAGGAATGTTATCAAGGAGTTCATGAGGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 ATGAGGTCTTCCAGGAATGTTATCAAGGAGTTCATGAGGTTTAAAGTTTCGCATGGAAGGA 60
Qy 61 ACGGTCAATGGGCAACGAGTTTGAATAAGAGGCGAAGGAGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGGCAACGAGTTTGAATAAGAGGCGAAGGAGGGAGGCCATACGAAGGC 120
Qy 121 CACATACCGTAAAGCTTAAAGTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCGGATATT 180
Db 121 CACATACCGTAAAGCTTAAAGTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCGGATATT 180
Qy 181 TTGTCCACCACAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTTCCCGCATACCA 240
Db 181 TTGTCCACCACAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTTCCCGCATACCA 240
Qy 241 GACTATATAAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATATAAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 301 GACGGTGGCGTGGTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTGGTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGTCAAGTTCATGGCGTGAACCTTCTTCCGATGGACCTGTATGCAAAAGAGACA 420
Db 361 AAGTCAAGTTCATGGCGTGAACCTTCTTCCGATGGACCTGTATGCAAAAGAGACA 420
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 480


```
PD 15-APR-2004.
XX
XX
XX 05-SEP-2003; 2003WO-US028078.
XX
XX 05-SEP-2002; 2002US-0408297P.
XX (VERT-) VERTEX PHARM INC.
XX
XX Whitney MA, Zeh K, Sanders PS;
XX
XX WPI; 2004-330208/30.
XX P-PSDB; ADM97769.
XX
XX Developing a sensor cell, useful in determining the activity of a target
XX gene and in developing therapeutic drugs, comprises providing cells
XX comprising a signal transduction detection system and introducing DNA
XX construct into cells.
XX
XX Disclosure; Page 167-168; 234pp; English.
XX
XX The present invention relates to a method of developing a sensor cell,
XX for determining the activity of a target gene in the cell, which
XX comprises providing a homogeneous population of cells, where each of the
XX cells comprises a signal transduction detection system and introducing
XX into the population of cells an isolated DNA construct comprising a
XX promoter operatively linked to a targeting sequence. The method is useful
XX in developing a sensor cell for determining the activity of a target gene
XX in the cell. The sensor cell and the methods are useful in developing new
XX and therapeutic drugs directed to the targets. The present sequence is a
XX coding sequence shown in the exemplification of the invention.
XX
XX Sequence 678 BP; 205 A; 129 C; 178 G; 166 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 12; Length 678;
Best Local Similarity 100.0%; Pred. No. 9.8e-208;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGCTCTCCAGAAGTGTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
DB 1 ATGAGGCTCTCCAGAAGTGTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
QY 61 ACGGTCAATGGCCAGCGAGTTTGAATAGAGCGGAAGGAGAGGGAGGCCATACGAAGGC 120
DB 61 ACGGTCAATGGCCAGCGAGTTTGAATAGAGCGGAAGGAGAGGGAGGCCATACGAAGGC 120
QY 121 CACATACCGTTAAGCTTAAGGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
DB 121 CACATACCGTTAAGCTTAAGGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTGCACATACCA 240
DB 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTGCACATACCA 240
QY 241 GACTATAAAAAGCTGTCAATTTCTGAAGGATTTTAAATGGGAAGGGTTCATGAACCTTTGAA 300
DB 241 GACTATAAAAAGCTGTCAATTTCTGAAGGATTTTAAATGGGAAGGGTTCATGAACCTTTGAA 300
QY 301 GACGGTGGCGTCTTACTGTACCCAGGATTCAGTTTGCAGGATGGCTTTTCATCTAC 360
DB 301 GACGGTGGCGTCTTACTGTACCCAGGATTCAGTTTGCAGGATGGCTTTTCATCTAC 360
QY 361 AAGGTCGAAGTTTCATTGGGCGTAACCTTTCTCCGATGGACCTGTATGCAAAAAGAGACA 420
DB 361 AAGGTCGAAGTTTCATTGGGCGTAACCTTTCTCCGATGGACCTGTATGCAAAAAGAGACA 420
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGGCGTGTCAAAAGGAGAG 480
DB 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGGCGTGTCAAAAGGAGAG 480
QY 481 ATTCATAGGCTCTGAAGCTGAAGACCGTGGTCAATTACCTAGTTGATTTCAAAAGTATT 540
DB 481 ATTCATAGGCTCTGAAGCTGAAGACCGTGGTCAATTACCTAGTTGATTTCAAAAGTATT 540

us-10-006-922a-11.rng
QY 541 TACATGGCAAGAGAGCGCTGTGACGTACCAAGGGTACTACTATGTTGACTCCAAACTGGAT 600
DB 541 TACATGGCAAGAGAGCGCTGTGACGTACCAAGGGTACTACTATGTTGACTCCAAACTGGAT 600
QY 601 ATAAAGCCACCAACGAAGACTATACAATCGTTGAGCAGTATGAAGAACCAGGAGGACGC 660
DB 601 ATAAAGCCACCAACGAAGACTATACAATCGTTGAGCAGTATGAAGAACCAGGAGGACGC 660
QY 661 CACCAATCTGTTCTCTTTAA 678
DB 661 CACCAATCTGTTCTCTTTAA 678

RESULT 10
AAL47952
ID AAL47952 standard; DNA; 859 BP.
XX
XX AAL47952;
XX
XX 26-SEP-2002 (first entry)
XX
XX Discosoma red fluorescent protein coding sequence.
XX
XX Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
XX modified yeast strain; environmental pollution; gene; ds.
XX
XX Discosoma sp.
XX
XX Key Location/Qualifiers
XX CDS 54..731
XX /*tag= a
XX /product= "red fluorescent protein"
XX
XX DE10061872-Al.
XX
XX 20-JUN-2002.
XX
XX 12-DEC-2000; 2000DB-01061872.
XX
XX 12-DEC-2000; 2000DB-01061872.
XX
XX (LICH/) LICHTENBERG-FRATE H.
XX
XX Lichtenberg-Frate H;
XX
XX WPI; 2002-539633/58.
XX P-PSDB; AAO18270.
XX
XX Modified yeast strain, useful for detecting toxic compounds in
XX environment, contains integrated cassettes responsive to genotoxic and
XX cytotoxic compounds.
XX
XX Disclosure; Page 20-21; 34pp; German.
XX
XX The present invention relates to a modified yeast strain that contains,
XX integrated stably and functionally in its genome, a genotoxicity cassette
XX and a cytotoxicity cassette, each comprising a promoter and reporter
XX gene, both of which are different in the two cassettes. The modified
XX yeast strain is used to detect environmental pollution, especially
XX genotoxic and/or cytotoxic substances in complex environmental
XX contaminants, especially organic compounds, but also (non-)ionising
XX radiation and chemical carcinogens. Particular applications are in
XX monitoring (waste) water (e.g. as an early warning system), medical
XX toxicology screening and for industrial process control. The present
XX sequence is a marker gene suitable for use in the cassettes of the
XX present invention
XX
XX Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;

Query Match 100.0%; Score 678; DB 6; Length 859;
Best Local Similarity 100.0%; Pred. No. 1.1e-207;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGAGGTCTTCCAAAGATTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATCGAAGGA 60
Db 54 ATGAGGTCTTCCAAAGATTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATCGAAGGA 113
QY 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGGAGGAGGAGGCCATACGAAGGC 120
Db 114 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGGAGGAGGAGGCCATACGAAGGC 173
QY 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGGCATTTGCTTGGGATATT 180
Db 174 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGGCATTTGCTTGGGATATT 233
QY 181 TTGTCAACACAAATTTCAATGAGGAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 240
Db 234 TTGTCAACACAAATTTCAATGAGGAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 293
QY 241 GACTATATAAAGCTGTCTCAATTTCTCAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 300
Db 294 GACTATATAAAGCTGTCTCAATTTCTCAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 353
QY 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTTTTCATCTAC 360
Db 354 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTTTTCATCTAC 413
QY 361 AAGGTCAAGTTCANTGGCGTGAACCTTTCTCCGATGGACCTGTATGCAAAAGAGACA 420
Db 414 AAGGTCAAGTTCANTGGCGTGAACCTTTCTCCGATGGACCTGTATGCAAAAGAGACA 473
QY 421 ATGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGGCGTTCGAAAGGAGAG 480
Db 474 ATGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGGCGTTCGAAAGGAGAG 533
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATACCTAGTTGAATTCAAAGTATT 540
Db 534 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATACCTAGTTGAATTCAAAGTATT 593
QY 541 TACATGGCAAGAGCCGTGTCAGCTACCGAGTACTACTATGTTGACTCCCAACTGGAT 600
Db 594 TACATGGCAAGAGCCGTGTCAGCTACCGAGTACTACTATGTTGACTCCCAACTGGAT 653
QY 601 ATAAAGCCCAACAGAGAGCTATACAACTCGTTCAGCAGTATGAAGAACCGAGGAGCG 660
Db 654 ATAAAGCCCAACAGAGAGCTATACAACTCGTTCAGCAGTATGAAGAACCGAGGAGCG 713
QY 661 CACCATCTGTTCCCTTTAA 678
Db 714 CACCATCTGTTCCCTTTAA 731
```

RESULT 11

```
AAD53432
ID AAD53432 standard; DNA; 859 BP.
XX AC AAD53432;
XX AC AAD53432;
DT 28-MAY-2003 (first entry)
XX DE
XX DE Discosoma species red fluorescent protein (RFP) encoding DNA.
XX KW Phosphorylation indicator; fluorescent protein; detection; phosphatase;
XX KW kinase; red fluorescent protein; RFP; gene; ds.
XX OS Discosoma sp.
XX FH Key Location/Qualifiers
XX CDS 54..731
XX FT /*tag= a
XX FT /product= "Discosoma sp. red fluorescent protein (RFP)"
XX PN WO200295058-A2.
XX PD 28-NOV-2002.
XX XX
```

```
PF 24-MAY-2002; 2002WO-US016955.
XX 24-MAY-2001; 2001US-00865291.
XX (REGC ) UNIV CALIFORNIA.
XX Tsien RY, Ting AY, Zhang J;
XX WPI; 2003-148474/14.
DR P-PSDB; AAE34962.
XX Novel chimeric phosphorylation indicators, useful for detecting
PT kinase/phosphatase in samples, has donor molecule, phosphorylatable
PT domain, phosphoaminoacid binding domain, and acceptor molecule, in
PT operative linkage.
XX Disclosure; Col 64-65; 38pp; English.
XX The present invention relates to chimeric phosphorylation indicators
CC comprising a phosphorylation polypeptide and a fluorescent protein or in
CC operative linkage, a donor molecule, a phosphorylatable domain, a
CC phosphoaminoacid binding domain (PAAABD) and an acceptor molecule. The
CC phosphorylation indicators of the invention are useful for detecting
CC kinases or phosphatases in a biological sample. They are also useful in
CC high throughput analysis e.g. for detecting a kinase inhibitor or
CC phosphatase inhibitor. The present sequence is Discosoma species red
CC fluorescent protein (RFP) encoding DNA used in the invention
XX SQ Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;
```

```
Query Match 100.0%; Score 678; DB 8; Length 859;
Best Local Similarity 100.0%; Pred. No. 1.1e-207;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGGTCTTCCAAAGATTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATCGAAGGA 60
Db 54 ATGAGGTCTTCCAAAGATTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATCGAAGGA 113
QY 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGGAGGAGGCCATACGAAGGC 120
Db 114 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGGAGGAGGCCATACGAAGGC 173
QY 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGGCATTTGCTTGGGATATT 180
Db 174 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGGCATTTGCTTGGGATATT 233
QY 181 TTGTCAACACAAATTTCAATGAGTATGAAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 240
Db 234 TTGTCAACACAAATTTCAATGAGTATGAAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 293
QY 241 GACTATATAAAGCTGTCTCAATTTCTCAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 300
Db 294 GACTATATAAAGCTGTCTCAATTTCTCAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 353
QY 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTTTTCATCTAC 360
Db 354 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTTTTCATCTAC 413
QY 361 AAGGTCAAGTTCANTGGCGTGAACCTTTCTCCGATGGACCTGTATGCAAAAGAGACA 420
Db 414 AAGGTCAAGTTCANTGGCGTGAACCTTTCTCCGATGGACCTGTATGCAAAAGAGACA 473
QY 421 ATGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGGCGTTCGAAAGGAGAG 480
Db 474 ATGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCGTATGGCGTTCGAAAGGAGAG 533
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATACCTAGTTGAATTCAAAGTATT 540
Db 534 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATACCTAGTTGAATTCAAAGTATT 593
QY 541 TACATGGCAAGAGCCGTGTCAGCTACCGAGTACTACTATGTTGACTCCCAACTGGAT 600
Db 594 TACATGGCAAGAGCCGTGTCAGCTACCGAGTACTACTATGTTGACTCCCAACTGGAT 653
```

```
QY 601 ATAACAAGCCACAACGAGCTATATACATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
D 114 ACGTCAATGGGCACGAGTTTGAATAGAAAGCGAAGGAGGGAGGCCATACGAGGC 173
Db 654 ATAACAAGCCACAACGAGCTATATACATCGTTGAGCAGTATGAAGAACCGAGGGACGC 713
QY 661 CACCATCTGTTCTCTTTAA 678
D 714 CACCATCTGTTCTCTTTAA 731
Db 714 CACCATCTGTTCTCTTTAA 731

RESULT 12
AAD61969
ID AAD61969 standard; cDNA; 859 BP.
XX AC AAD61969;
XX 15-JAN-2004 (first entry)
DE Discosoma sp. red fluorescent protein (RFP) cDNA.
KW Fluorescent protein; resonance energy transfer; pH; detection;
KW red fluorescent protein; RFP; gene; ss.
XX OS
XX TSien RY, Zacharias DA, Baird GS;
XX Key Location/Qualifiers
XX CDS 54..731
XX /*tag= a
XX /product= "Discosoma sp. red fluorescent protein"
XX US2003170911-A1.
XX 11-SEP-2003.
XX 26-FEB-2001; 2001US-00794308.
XX 26-FEB-2001; 2001US-00794308.
XX (TSIE/) TSien R Y.
XX (ZACH/) ZACHARIAS D A.
XX (BAIR/) BAIRD G S.
XX TSien RY, Zacharias DA, Baird GS;
XX WPI; 2003-802418/75.
XX P-PSDB; ABW00918.
XX Fluorescent proteins containing a mutation that reduces or eliminates its
XX ability to oligomerize which gives more reliable fluorescence resonance
XX energy transfer results and are useful to detect molecule interaction,
XX enzymes, or sample pH.
XX Disclosure; Page 29-30; Opp: English.
XX The invention relates to a non-oligomerising fluorescent protein
XX containing a mutation that reduces or eliminates its ability to
XX oligomerise. The fluorescent protein gives more reliable fluorescence
XX resonance energy transfer results and are useful to detect molecule
XX interaction, enzymes, or sample pH. These are also used to identify
XX agents or conditions that regulate expression of control sequences. The
XX present sequence is Discosoma sp. red fluorescent protein (RFP) cDNA
XX
XX Sequence 859 BP; 263 A; 163 C; 228 G; 205 T; 0 U; 0 Other;
XX
XX Query Match
XX Best Local Similarity 100.0%; Score 678; DB 10; Length 859;
XX Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 ATGAGGTTCTCCACGATGTTATCAAGAGTTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
D 54 ATGAGGTTCTCCACGATGTTATCAAGAGTTCATGAGTTTAAAGTTTCGCATGGAAGGA 113
QY 61 ACGTCAATGGGCACGAGTTTGAATAGAAAGCGAAGGAGGGAGGCCATACGAAAGGC 120
```

```
RESULT 13
ADF70404
ID ADF70404 standard; cDNA; 678 BP.
XX AC ADF70404;
XX 12-FEB-2004 (first entry)
XX Discosoma wild-type GFP variant cDNA SeqID27.
DE ligand; orphan receptor protein; fusion protein; fluorescent protein;
DE cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
DE GFPuv; Enhanced GFP; EGFP; gene; ss.
XX OS Discosoma sp.
XX WO2003071272-A1.
XX 28-AUG-2003.
XX 21-FEB-2003; 2003WO-JP001901.
XX 22-FEB-2002; 2002JP-00045728.
XX 23-JUL-2002; 2002JP-00213949.
XX 11-OCT-2002; 2002JP-00298237.
XX (TAKE ) TAKEDA CHEM IND LTD.
XX Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
```



```
XX WPI; 2003-697654/66.
DR P-PSDB; ADF70403.
XX
PT Transformation of cells with a fusion protein of an orphan receptor
PT protein with a fluorescent protein useful for identification of ligands
PT to the orphan receptor.
XX
XX Disclosure; SEQ ID NO 27; 594pp; Japanese.
XX
CC This invention relates to a novel method of identifying ligands to an
CC orphan receptor protein which comprises transforming cells with DNA
CC encoding a fusion protein of the orphan receptor with a fluorescent
CC protein, so that the fusion protein is expressed in the cells (or cell
CC membranes isolated from them) and contacting the cells with the potential
CC ligand to be tested. A suitable fluorescent protein for incorporation in
CC the fusion protein is green fluorescent protein (GFP), for example GFP-1,
CC wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
CC identification of ligands binding to an orphan receptor protein.
XX
XX Sequence 678 BP; 204 A; 129 C; 179 G; 166 T; 0 U; 0 Other;
XX
XX Query Match 99.9%; Score 677; DB 10; Length 678;
XX Best Local Similarity 100.0%; Pred. No. 2.1e-207;
XX Matches 677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTCATGAGTTTAAGGTTTCGATGGAAGGA 60
Db 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTCATGAGTTTAAGGTTTCGATGGAAGGA 60
Qy 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGGGGAGGCCATACGAAGGC 120
Qy 121 CACATACCGTAAAGCTTAAGCTAACCAAGGGGAGCCCTTGGCCATTGCTTGGGATATT 180
Db 121 CACATACCGTAAAGCTTAAGCTAACCAAGGGGAGCCCTTGGCCATTGCTTGGGATATT 180
Qy 181 TTGTACACCAATTTCAATGATGGAAGCAAGGTATATGTCAAGCAACCTGCCGACATACCA 240
Db 181 TTGTACACCAATTTCAATGATGGAAGCAAGGTATATGTCAAGCAACCTGCCGACATACCA 240
Qy 241 GACTATAAAGCTGTCATTTCTGAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATAAAGCTGTCATTTCTGAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGTCAAGTTTCATGGCGTGAACCTTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGTCAAGTTTCATGGCGTGAACCTTTCTCCGATGACCTGTTATGCAAAAGAGACA 420
Qy 421 ATGGGCTGGGAAGCAGCAGCTGAGCTTTGATCCTCGTGTATGGCGTGTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCAGCAGCTGAGCTTTGATCCTCGTGTATGGCGTGTGAAAGGAGAG 480
Qy 481 ATTCATAGGCTCTGAAGCTCAAGAGCGTGGTCAATACCTAGTTGAATTCAAAGTATT 540
Db 481 ATTCATAGGCTCTGAAGCTCAAGAGCGTGGTCAATACCTAGTTGAATTCAAAGTATT 540
Qy 541 TACATGGCAAAAGAGCCCTGTGACGTACCCAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAAAGAGCCCTGTGACGTACCCAGGCTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATACAGCCCAACAGGAAGCTATACATCTGTTGAGCAGGTATGAAGAACCGAGGGACGC 660
Db 601 ATACAGCCCAACAGGAAGCTATACATCTGTTGAGCAGGTATGAAGAACCGAGGGACGC 660
Qy 661 CACCATCTGTTCCCTTTA 677
Db 661 CACCATCTGTTCCCTTTA 677
```

RESULT 14

```
ACA62995
XX ACA62995 standard; DNA; 3311 BP.
AC ACA62995;
XX
XX 23-SEP-2003 (first entry)
XX
XX Plasmid DNA containing coding sequence for RFP.
XX
XX Fluorescently-tagged enzyme; substrate; cell population; GFP;
XX quantification of enzymatic activity; green fluorescent protein;
XX red fluorescent protein; RFP; enhanced green fluorescent protein; EGFP;
XX enzymatic process; cellular enzyme; chemotherapeutic drug;
XX multidrug resistance; MDR; ds.
XX
XX Aequorea victoria.
XX Synthetic.
XX
XX US2003049597-A1.
XX
XX 13-MAR-2003.
XX
XX 01-MAR-2001; 2001US-00797496.
XX
XX 01-MAR-2001; 2001US-00797496.
XX
XX (SIMO/) SIMON S M.
XX (CHEN/) CHEN Y.
XX
XX Simon SM, Chen Y;
XX
XX WPI; 2003-555145/52.
XX
XX Simultaneously quantifying in situ the relationship between an enzyme and
XX its substrate for study of enzymatic processes at a cellular level,
XX comprises optical measurements on cells expressing the fluorescently
XX tagged enzyme.
XX
XX Disclosure; Page 8-9; 21pp; English.
XX
XX The present invention relates to a method for simultaneously quantifying
XX in situ the relationship between a fluorescently-tagged enzyme (E) and
XX its substrate in a population of cells. A population of cells expressing
XX (E) is created in which different cells contain different amounts of (E).
XX The cells are incubated with a substrate, and the concentration and
XX enzymatic activity of (E) in each cell are simultaneously quantified by
XX optical means. The fluorescently-tagged enzyme is produced by linking the
XX enzyme with Aequorea victoria green fluorescent protein (GFP), or its
XX variants or derivatives such as red fluorescent protein (RFP) and
XX enhanced green fluorescent protein (EGFP). The method of the invention is
XX useful for simultaneously quantifying in situ the relationship between an
XX enzyme and its substrate in a cell. The method is useful for the study of
XX enzymatic processes at the cellular level, and especially for examining
XX the interactions of cellular enzymes with chemotherapeutic drugs, e.g.
XX for elucidating mechanisms of multidrug resistance (MDR). The present
XX sequence represents a plasmid containing the coding sequence for RFP
XX
XX Sequence 3311 BP; 858 A; 778 C; 860 G; 815 T; 0 U; 0 Other;
XX
XX Query Match 99.9%; Score 677; DB 9; Length 3311;
XX Best Local Similarity 100.0%; Pred. No. 4.6e-207;
XX Matches 677; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
Qy 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTCATGAGTTTAAGGTTTCGATGGAAGGA 60
Db 289 ATGAGGCTTCCAGAAATGTTATCAAGAGTTCATGAGTTTAAGGTTTCGATGGAAGGA 348
Qy 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGGGGAGGCCATACGAAGGC 120
Db 349 ACGGTCAATGGGACGAGTTTGAATAGAGGCGAAGGAGGGGAGGCCATACGAAGGC 408
```

Qy 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGAGCCTTTGGCCATTTGCTTGGATATT 180
Db |||||
Qy 409 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGAGCCTTTGGCCATTTGCTTGGATATT 468
Db |||||
Qy 181 TTGTCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240
Db TTGTCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 528
Qy 241 GACTATAAAAGCTGTCTATTTCTGGAAGGATTTAAATGGGAAAGGTCATGAATTTGAA 300
Db GACTATAAAAGCTGTCTATTTCTGGAAGGATTTAAATGGGAAAGGTCATGAATTTGAA 588
Qy 301 GACGGTGGCGTCTGACTTACTTAACCCAGGATTTCCAGTTTGCAGGATGGCTTTTCATCTAC 360
Db GACGGTGGCGTCTGACTTACTTAACCCAGGATTTCCAGTTTGCAGGATGGCTTTTCATCTAC 648
Qy 361 AAGGTCAAGTTTCATTTGGCGTGAACCTTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db AAGGTCAAGTTTCATTTGGCGTGAACCTTTCCGATGGACCTGTTATGCAAAAGAGACA 708
Qy 421 ATGGCTGGGAAGCCAGCACTAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 480
Db ATGGCTGGGAAGCCAGCACTAGCGTTTGTATCTCTGTATGGCGTGTGAAAGGAGAG 768
Qy 481 ATTCAATAGGCTCTGAAGCTGAAGAGACGCTGCTCATTTACCTAGTTGAATTCAAAAGTATT 540
Db ATTCAATAGGCTCTGAAGCTGAAGAGACGCTGCTCATTTACCTAGTTGAATTCAAAAGTATT 828
Qy 541 TACATGGCAAAAGAGCCCTGTGAGCTACAGGCTACTACTATGTTGAATCCTCAAACTGGAT 600
Db TACATGGCAAAAGAGCCCTGTGAGCTACAGGCTACTACTATGTTGAATCCTCAAACTGGAT 888
Qy 601 ATAACAGCCCAACAGGAGACTATACAACTGTTGAGCAGATATGAAGAACCGAGGGACGC 660
Db ATAACAGCCCAACAGGAGACTATACAACTGTTGAGCAGATATGAAGAACCGAGGGACGC 948
Qy 661 CACCATCTGTTCCCTTTA 677
Db CACCATCTGTTCCCTTTA 965
RESULT 15
ABL61142
ID ABL61142 standard; DNA; 666 BP.
AC ABL61142;
XX
DT 18-SEP-2002 (first entry)
XX Red fluorescent protein reporter DNA fragment.
XX Alpha-SMA; smooth muscle actin; screening; smooth muscle cell;
KW myofibroblast; gene therapy; red fluorescent protein; ds.
XX Synthetic.
XX
PN EP1172375-A1.
XX
PD 16-JAN-2002.
XX
PF 22-DEC-2000; 2000EP-00128446.
XX
PR 11-JUL-2000; 2000DE-01033633.
PR 31-OCT-2000; 2000DE-01053879.
XX
PA (ODEN/) ODENTHAL M.
XX
PI Odenthal M, Jung D;
XX
DR WPI; 2002-149590/20.
XX
PT New nucleic acid containing regulatory region of the smooth muscle actin

PT Gene, useful e.g. for manipulating gene expression in smooth muscle cells.
XX
PS Disclosure; Page 20; 44pp; German.
XX
CC This invention describes a novel nucleic acid (I) comprising: (i) at least one functional region (Ia) from the regulatory region of the alpha-smooth muscle actin (SMA) gene and (ii) at least one additional functional sequence (Ib) operably linked to (Ia). The products of the invention can be used for preparing genetically modified eukaryotic cells or organisms, for isolation and screening of smooth muscle cells, myofibroblasts or related cells, and for manipulation of gene expression and/or cell function in smooth muscle cell or myofibroblasts, particularly for gene therapy. Component (Ia) provides cell-type- or differentiation-specific expression or modulation of genes. This sequence represents a DNA fragment of red fluorescent protein (RFP) which can be used as a reporter molecule under the control of the alpha-SMA described in the invention
XX
SQ Sequence 666 BP; 202 A; 126 C; 176 G; 162 T; 0 U; 0 Other;
Query Match 98.1%; Score 665; DB 6; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.5e-203;
Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 13 AAGAATGTTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGCAAGTCAATGGG 72
Db 1 AAGAATGTTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGCAAGTCAATGGG 60
Qy 73 CACGAGTTTGAATAGAACGCGAAGGAGGAGGCGGCATACGAAGGCCCAATATACCGTA 132
Db 61 CACGAGTTTGAATAGAACGCGAAGGAGGAGGCGGCATACGAAGGCCCAATATACCGTA 120
Qy 133 AAGCTTAAGTAAACCAAGGGGAGCCTTTGCCATTTGCTTGGGATATTTGTCAACCAAA 192
Db 121 AAGCTTAAGTAAACCAAGGGGAGCCTTTGCCATTTGCTTGGGATATTTGTCAACCAAA 180
Qy 193 TTTTCAGTATGGAAGCAAGGTATATGCAAGCACCTCGCCGACATACCAGACTATAAAAG 252
Db 181 TTTTCAGTATGGAAGCAAGGTATATGCAAGCACCTCGCCGACATACCAGACTATAAAAG 240
Qy 253 CTGTCATTTCTGAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAAGACGTTGGCGTC 312
Db 241 CTGTCATTTCTGAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAAGACGTTGGCGTC 300
Qy 313 GTTACTGTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTTCATCTACAGGTCNAGTTC 372
Db 301 GTTACTGTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTTCATCTACAGGTCNAGTTC 360
Qy 373 ATTGGCGTGAACCTTTCTCTCCGATGGACCTGTTATGCAAAAGAGACAATGGGCTGGAA 432
Db 361 ATTGGCGTGAACCTTTCTCTCCGATGGACCTGTTATGCAAAAGAGACAATGGGCTGGAA 420
Qy 433 GCCAGCACTGAGCGTTTGTATCTCTGATGAGCGTGTGTAAGAGGAGATTCATTAAGGCT 492
Db 421 GCCAGCACTGAGCGTTTGTATCTCTGATGAGCGTGTGTAAGAGGAGATTCATTAAGGCT 480
Qy 493 CTGAAGCTGAAGACGCTGTCATTTACCTAGTTGAATTCAAAAGTATTTACATGGCAGAG 552
Db 481 CTGAAGCTGAAGACGCTGTCATTTACCTAGTTGAATTCAAAAGTATTTACATGGCAGAG 540
Qy 553 AAGCTGTGCAGCTTACCAGGGTACTACTATGTTGACTCCAAACTGGATATAAACAAGCCAC 612
Db 541 AAGCTGTGCAGCTTACCAGGGTACTACTATGTTGACTCCAAACTGGATATAAACAAGCCAC 600
Qy 613 AACGAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGAGCGCCACCATCTGTTTC 672
Db 601 AACGAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGAGCGCCACCATCTGTTTC 660
Qy 673 CTTTA 677
Db 661 CTTTA 665

Search completed: July 1, 2005, 16:33:30
Job time : 1033 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2005, 16:16:18 ; Search time 571 Seconds
(without alignments)
7447.532 Million cell updates/sec

Title: US-10-006-922a-11

Perfect score: 678

Sequence: 1 ataggtttccaaagtgt.....gccaccatctgttcctttaa 678

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 6313374 seqs, 3136092125 residues

Total number of hits satisfying chosen parameters: 12626748

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq.*
19: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq.*
20: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq.*
21: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq.*
22: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
23: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq.*
24: /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq.*
25: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
26: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	678	100.0	678	9	US-09-967-772-6
2	678	100.0	678	13	US-10-006-922-11
3	678	100.0	678	14	US-10-121-258-2
4	678	100.0	678	17	US-10-132-067-3
5	678	100.0	678	17	US-10-335-517-6
6	678	100.0	678	17	US-10-334-288-6
7	678	100.0	678	19	US-10-311-030-5

Query Match 100.0%; Score 678; DB 9; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;

8	678	100.0	678	21	US-10-656-029-21	Sequence 21, Appl
9	678	100.0	859	9	US-09-999-745-66	Sequence 66, Appl
10	678	100.0	859	10	US-09-866-538-11	Sequence 11, Appl
11	678	100.0	859	10	US-09-794-308-11	Sequence 11, Appl
12	678	100.0	859	10	US-09-865-291-11	Sequence 11, Appl
13	678	100.0	859	19	US-10-433-640-12	Sequence 12, Appl
14	678	100.0	859	20	US-10-885-988-11	Sequence 11, Appl
15	678	100.0	859	21	US-10-857-622-11	Sequence 11, Appl
16	677	99.9	3311	10	US-10-505-486-27	Sequence 27, Appl
17	677	99.9	3311	10	US-09-797-496B-3	Sequence 3, Appl
18	665	98.1	666	19	US-10-332-733-22	Sequence 22, Appl
19	661	97.5	711	19	US-10-314-936-1	Sequence 1, Appl
20	661	97.5	711	19	US-10-314-936-3	Sequence 3, Appl
21	661	97.5	711	23	US-11-021-014-1	Sequence 1, Appl
22	661	97.5	711	23	US-11-021-014-3	Sequence 3, Appl
23	658.8	97.2	898	13	US-10-006-922-45	Sequence 45, Appl
24	642.8	94.8	678	18	US-10-423-688A-40	Sequence 40, Appl
25	566.4	83.5	876	13	US-10-006-922-17	Sequence 17, Appl
26	566.4	83.5	876	15	US-10-161-403-39	Sequence 39, Appl
27	450.6	66.5	723	14	US-10-152-296-1	Sequence 1, Appl
28	450.6	66.5	723	19	US-10-739-656-1	Sequence 1, Appl
29	449.8	66.3	681	19	US-10-311-030-8	Sequence 8, Appl
30	449.8	66.3	713	19	US-10-311-030-11	Sequence 11, Appl
31	449.8	66.3	713	19	US-10-311-030-12	Sequence 12, Appl
32	449.2	66.3	678	14	US-10-081-864-7	Sequence 7, Appl
33	449.2	66.3	678	16	US-10-315-920-1	Sequence 1, Appl
34	447.2	66.0	681	14	US-10-121-258-3	Sequence 3, Appl
35	447.2	66.0	681	14	US-10-121-258-23	Sequence 23, Appl
36	447.2	66.0	4692	15	US-10-161-403-29	Sequence 16, Appl
37	447.2	66.0	4692	19	US-10-433-640-16	Sequence 45, Appl
38	447.2	66.0	6984	13	US-10-001-189-45	Sequence 45, Appl
39	447.2	66.0	7910	21	US-10-169-050-20	Sequence 20, Appl
40	447.2	66.0	9320	19	US-10-471-065-20	Sequence 4, Appl
41	447.2	66.0	9658	19	US-10-609-019-4	Sequence 3, Appl
42	447.2	66.0	9678	19	US-10-609-019-3	Sequence 2, Appl
43	447.2	66.0	10263	19	US-10-609-019-2	Sequence 51, Appl
44	446	65.8	678	16	US-10-315-920-3	
45	445.8	65.8	1638	15	US-10-214-932-51	

ALIGNMENTS

RESULT 1

US-09-967-772-6
; Sequence 6, Application US/0996772
; Patent No. US20020164577A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TS'EN, Roger
; APPLICANT: GONZALEZ, Jesus

; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/09/967,772
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US 09/459,956
; PRIOR FILING DATE: 1999-12-13

; PRIOR APPLICATION NUMBER: US 08/765,860
; PRIOR FILING DATE: 1996-12-19
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: US 08/481,977
; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA

; ORGANISM: Discosoma sp "red"
US-09-967-772-6

	Matches	678;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	ATGAGGTC	TCCAAAGA	TGTTATCAAGGAG	TTTCATGAGGTTTAAAGGTTTCGATCGAAGGA	60				
Db	1	ATGAGGTC	TCCAAAGA	TGTTATCAAGGAG	TTTCATGAGGTTTAAAGGTTTCGATCGAAGGA	60				
Qy	61	ACGGTCAAT	GGGCGACGAG	TTTGAATAGAACGG	CGAAGAGAGGGGAGGCCATACGAAGGC	120				
Db	61	ACGGTCAAT	GGGCGACGAG	TTTGAATAGAACGG	CGAAGAGAGGGGAGGCCATACGAAGGC	120				
Qy	121	CACAATACC	GTAAAGCTTAAG	TAAACCAAGGGGGAC	CTTTGCCATTGTCTTGGGATATT	180				
Db	121	CACAATACC	GTAAAGCTTAAG	TAAACCAAGGGGGAC	CTTTGCCATTGTCTTGGGATATT	180				
Qy	181	TTGTCA	CCACAATTTTCAG	TATGGAAAGCAAGGTATATGTCAAGCA	CCCTTCCGCGCATACCA	240				
Db	181	TTGTCA	CCACAATTTTCAG	TATGGAAAGCAAGGTATATGTCAAGCA	CCCTTCCGCGCATACCA	240				
Qy	241	GACTATAAA	AAGCTGTCA	TTTCTGAAAGGATTTAAATGGG	AAAGGGTCATGAACTTTGAA	300				
Db	241	GACTATAAA	AAGCTGTCA	TTTCTGAAAGGATTTAAATGGG	AAAGGGTCATGAACTTTGAA	300				
Qy	301	GACGGTGG	CGTCGTACTGT	TAAACCCAGGATTCAG	TGTTGCAAGATGGCTTTTCATCTAC	360				
Db	301	GACGGTGG	CGTCGTACTGT	TAAACCCAGGATTCAG	TGTTGCAAGATGGCTTTTCATCTAC	360				
Qy	361	AAGSTCAAG	TTTCATTTGGCGT	GAACTTTTCTTCCATG	SGACCTGTATGCAAAAGACACA	420				
Db	361	AAGSTCAAG	TTTCATTTGGCGT	GAACTTTTCTTCCATG	SGACCTGTATGCAAAAGACACA	420				
Qy	421	ATGGGCTGG	GAAGCAGC	ACTGAGCGTTTG	TATCCTCGTATGGCGTGTTCGAAAGGAGAG	480				
Db	421	ATGGGCTGG	GAAGCAGC	ACTGAGCGTTTG	TATCCTCGTATGGCGTGTTCGAAAGGAGAG	480				
Qy	481	ATTCAT	AAGGCTCTGAAG	CTGAAAGACGGTGGT	CATTACCTAGTTGAAATTCAAAAGTATT	540				
Db	481	ATTCAT	AAGGCTCTGAAG	CTGAAAGACGGTGGT	CATTACCTAGTTGAAATTCAAAAGTATT	540				
Qy	541	TACATGCA	AGAGACCTGTG	CGACTACCAAGG	TACTATGTGTGACTCCAACTCGAT	600				
Db	541	TACATGCA	AGAGACCTGTG	CGACTACCAAGG	TACTATGTGTGACTCCAACTCGAT	600				
Qy	601	ATAACAAG	CCACAACGAAG	ACTATACAATCCGT	TGAGCAGTATGAAAGAACCGAGGGACGC	660				
Db	601	ATAACAAG	CCACAACGAAG	ACTATACAATCCGT	TGAGCAGTATGAAAGAACCGAGGGACGC	660				
Qy	661	CACCAT	CTGTTCTTTAA			678				
Db	661	CACCAT	CTGTTCTTTAA			678				

RESULT 2
 US-10-006-922-11
 ; Sequence 11, Application US/10006922
 ; Publication No. US20020197676A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lukyanov, Sergey A
 ; APPLICANT: Fradkov, Arcady F.
 ; APPLICANT: Labas, Yulii A.
 ; APPLICANT: Matz, Mikhail V.
 ; APPLICANT: Tersikh, Alexey
 ; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
 ; TITLE OF INVENTION: Methods for Using the Same
 ; FILE REFERENCE: CLON-035CIP
 ; CURRENT APPLICATION NUMBER: US/10/006,922
 ; CURRENT FILING DATE: 2001-12-04
 ; PRIOR APPLICATION NUMBER: 09/120,330
 ; PRIOR FILING DATE: 1998-12-11
 ; PRIOR APPLICATION NUMBER: 09/457,898
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,144
 ; PRIOR FILING DATE: 1999-12-09
 ; PRIOR APPLICATION NUMBER: 09/458,477

```

; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma species
; US-10-006-922-11

Query Match      100.0%; Score 678; DB 13; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  ATGAGGTCTTCCAAAGATGTTATCAAGGAGTTCATGAGGTTTAAAGGTTTCGCATCGGAAGGA 60
Db      1  ATGAGGCTCTCCAAAGATGTTATCAAGGAGTTCATGAGGTTTAAAGGTTTCGCATCGGAAGGA 60

Qy     61  ACGGTCAATGGGCACAGATTGAAATAGAGGCGAAGGAGAGGGGAGGCATACGAGGC 120
Db     61  ACGGTCAATGGGCACAGATTGAAATAGAGGCGAAGGAGAGGGGAGGCATACGAGGC 120

Qy    121  CACAATACCGTAAAGCTTAAAGGTAAACAAGGGGGGACCTTTGGCCATTGTCTTGGGATATT 180
Db    121  CACAATACCGTAAAGCTTAAAGGTAAACAAGGGGGGACCTTTGGCCATTGTCTTGGGATATT 180

Qy    181  TTGTCAACCAAAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTGCGGCACATACCA 240
Db    181  TTGTCAACCAAAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTGCGGCACATACCA 240

Qy    241  GACTATAAAAGGTGTCATTTCTCGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db    241  GACTATAAAAGGTGTCATTTCTCGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300

Qy    301  GACGGTGGCGTCGTACTGTAAACCCAGGATTCAGATTTGCAGGATGCGTGTTCATCTAC 360
Db    301  GACGGTGGCGTCGTACTGTAAACCCAGGATTCAGATTTGCAGGATGCGTGTTCATCTAC 360

Qy    361  AAGGTCAAGTTCAATGGCGTGAACCTTTCCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db    361  AAGGTCAAGTTCAATGGCGTGAACCTTTCCCTCCGATGGACCTGTTATGCAAAAGAGACA 420

Qy    421  ATGGGTGGGAAGCCAGCACTGAGCGTTTGATTCCTCGTGATGCGCGTGTGAAAGAGAG 480
Db    421  ATGGGTGGGAAGCCAGCACTGAGCGTTTGATTCCTCGTGATGCGCGTGTGAAAGAGAG 480

Qy    481  ATTCATAAGGCTCTGAAGCTGAAGACGCTGGTGCATTAAGTCTGAAATCAAAAGTATT 540
Db    481  ATTCATAAGGCTCTGAAGCTGAAGACGCTGGTGCATTAAGTCTGAAATCAAAAGTATT 540

Qy    541  TACATGCGAAGAGCGCTGTGCAGTACACAGGGTACTACTATGCTTGACTCCAAACTGGAT 600
Db    541  TACATGCGAAGAGCGCTGTGCAGTACACAGGGTACTACTATGCTTGACTCCAAACTGGAT 600

Qy    601  ATAAACAGGCACCAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGGAGGACGC 660
Db    601  ATAAACAGGCACCAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGGAGGACGC 660

Qy    661  CACCATCTGTTCTTTAA 678
Db    661  CACCATCTGTTCTTTAA 678

```

RESULT 3
US-10-121-258-2
; Sequence 2, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert

;; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
;; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
;; FILE REFERENCE: UC083.1C22CP1

;; CURRENT APPLICATION NUMBER: US/10/121.258

;; CURRENT FILING DATE: 2002-04-10

;; PRIOR APPLICATION NUMBER: 09/794,308

;; PRIOR FILING DATE: 2001-02-26

;; PRIOR APPLICATION NUMBER: 09/866,538

;; PRIOR FILING DATE: 2001-05-24

;; NUMBER OF SEQ ID NOS: 78

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 2

;; LENGTH: 678

;; TYPE: DNA

;; ORGANISM: Discosoma sp.

;; FEATURE:

;; NAME/KEY: misc feature

;; LOCATION: (1)...(678)

;; OTHER INFORMATION: wild-type DsRed

US-10-121-258-2

Query Match 100.0%; Score 678; DB 14; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ATGAGGCTCTCCAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60
Db 1 ATGAGGCTCTCCAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60

Qy 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGAGGCCATACGAAGGC 120

Qy 121 CACAATACCGTAAAGCTTAAGGTTAACAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAGGTTAACAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACTTCCGACATACCA 240
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACTTCCGACATACCA 240

Qy 241 GACTATAAAGAGCTGCTCAATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATAAAGAGCTGCTCAATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300

Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 361 AAGGTCAAGTTTCATTTGCGGTGAACCTTTCTTCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCAAGTTTCATTTGCGGTGAACCTTTCTTCGATGGACCTGTTATGCAAAAGAGACA 420

Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGATCCTCGTATGGCGTTTCAAGAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGATCCTCGTATGGCGTTTCAAGAGGAGAG 480

Qy 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTGAATTCAAAGATATT 540
Db 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTGAATTCAAAGATATT 540

Qy 541 TACATGCAAGAGCCCTGTGACGCTACAGGGTACTATGTTGATCTCAAACTGGAT 600
Db 541 TACATGCAAGAGCCCTGTGACGCTACAGGGTACTATGTTGATCTCAAACTGGAT 600

Qy 541 TACATGCAAGAGCCCTGTGACGCTACAGGGTACTATGTTGATCTCAAACTGGAT 600
Db 541 TACATGCAAGAGCCCTGTGACGCTACAGGGTACTATGTTGATCTCAAACTGGAT 600

Qy 601 ATACAGCCACACGAGGACTATACAACTGTTGAGCAGTATGAAGAACCGAGGGAGCG 660
Db 601 ATACAGCCACACGAGGACTATACAACTGTTGAGCAGTATGAAGAACCGAGGGAGCG 660

Qy 661 CACCATCTGTTCTTTTAA 678
Db 661 CACCATCTGTTCTTTTAA 678
```

RESULT 4

US-10-132-067-3

;; Sequence 3, Application US/10132067

;; Publication No. US2003020335A1

;; GENERAL INFORMATION:

;; APPLICANT: Bradbury, Andrew

;; APPLICANT: Zeytun, Ahmet

;; APPLICANT: Waldo, Geoffrey

;; APPLICANT: The Regents of the University of California

;; TITLE OF INVENTION: Fluorescent Binding Ligands With Intrinsic

;; FILE REFERENCE: 021362-000600US

;; CURRENT APPLICATION NUMBER: US/10/132.067

;; CURRENT FILING DATE: 2002-04-24

;; NUMBER OF SEQ ID NOS: 48

;; SOFTWARE: PatentIn Ver. 2.1

;; SEQ ID NO 3

;; LENGTH: 678

;; TYPE: DNA

;; ORGANISM: Discosoma sp.

;; FEATURE:

;; OTHER INFORMATION: red fluorescent protein (dsRED)

;; NAME/KEY: CDS

;; LOCATION: (1)...(678)

;; OTHER INFORMATION: dsRED

US-10-132-067-3

Query Match 100.0%; Score 678; DB 17; Length 678;

Best Local Similarity 100.0%; Pred. No. 5.9e-216;

Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 ATGAGGCTCTCCAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60
Db 1 ATGAGGCTCTCCAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60

Qy 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGGACGAGTTTGAATAGAGGCGCAAGGAGGGAGGCCATACGAAGGC 120

Qy 121 CACAATACCGTAAAGCTTAAGGTTAACAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAGGTTAACAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180

Qy 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACTTCCGACATACCA 240
Db 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACTTCCGACATACCA 240

Qy 241 GACTATAAAGAGCTGCTCAATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATAAAGAGCTGCTCAATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300

Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360

Qy 361 AAGGTCAAGTTTCATTTGCGGTGAACCTTTCTTCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCAAGTTTCATTTGCGGTGAACCTTTCTTCGATGGACCTGTTATGCAAAAGAGACA 420

Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGATCCTCGTATGGCGTTTCAAGAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGATCCTCGTATGGCGTTTCAAGAGGAGAG 480

Qy 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTGAATTCAAAGATATT 540
Db 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGTGGTCAATTACCTAGTTGAATTCAAAGATATT 540

Qy 541 TACATGCAAGAGCCCTGTGACGCTACAGGGTACTATGTTGATCTCAAACTGGAT 600
Db 541 TACATGCAAGAGCCCTGTGACGCTACAGGGTACTATGTTGATCTCAAACTGGAT 600
```

Hand date

QY 601 ATACCAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGACGC 660
Db 601 ATACCAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGACGC 660
QY 661 CACCATCTGTTCTTTAA 678
Db 661 CACCATCTGTTCTTTAA 678
RESULT 5
US-10-335-517-6
; Sequence 6, Application US/10335517
; Publication No. US20030207248A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/10/335,517
; CURRENT FILING DATE: 2002-12-31
; PRIOR FILING DATE: 2001-09-28
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1996-12-19
; PRIOR FILING DATE: 1996-06-06
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
US-10-335-517-6
Query Match 100.0%; Score 678; DB 17; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGGCTTCCCAAGATGTTATCAAGGAGTTCAAGGTTTAAAGGTTTCGATCGAAGGA 60
Db 1 ATGAGGCTTCCCAAGATGTTATCAAGGAGTTCAAGGTTTAAAGGTTTCGATCGAAGGA 60
QY 61 ACGGTCAATGGCCACGAGTTTGAATAGAGGCGAAGAGAGAGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGCCACGAGTTTGAATAGAGGCGAAGAGAGAGGAGGCCATACGAAGGC 120
QY 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGACCCCTGCCGACATACCA 240
Db 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGACCCCTGCCGACATACCA 240
QY 241 GACTATATAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATATAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTATGCGGTGTTGAAAGGAGAG 480

Db 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGTATGCGGTGTTGAAAGGAGAG 480
QY 481 ATTCTAAGGCTCTGAAGCTGGAAGACGGTGGTCAATTACCTAGTGTGAATTCAAAAGTATT 540
Db 481 ATTCTAAGGCTCTGAAGCTGGAAGACGGTGGTCAATTACCTAGTGTGAATTCAAAAGTATT 540
QY 541 TACATGGCAAGAGGCTGTGACGCTTACCAAGGCTACTTACTTATCTTGAATTCAAAAGTATT 600
Db 541 TACATGGCAAGAGGCTGTGACGCTTACCAAGGCTACTTACTTATCTTGAATTCAAAAGTATT 600
QY 601 ATACCAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGACGC 660
Db 601 ATACCAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAACCGAGGACGC 660
QY 661 CACCATCTGTTCTTTAA 678
Db 661 CACCATCTGTTCTTTAA 678
RESULT 6
US-10-334-288-6
; Sequence 6, Application US/10334288
; Publication No. US20040002123A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: GONZALEZ, Jesus
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY OPTICAL METHODS
; FILE REFERENCE: REGEN1290-5
; CURRENT APPLICATION NUMBER: US/10/334,288
; CURRENT FILING DATE: 2002-12-31
; PRIOR FILING DATE: 2001-09-28
; PRIOR FILING DATE: 2001-09-28
; PRIOR FILING DATE: 1999-12-13
; PRIOR FILING DATE: 1996-12-19
; PRIOR FILING DATE: 1996-06-06
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 678
; TYPE: DNA
; ORGANISM: Discosoma sp "red"
US-10-334-288-6
Query Match 100.0%; Score 678; DB 17; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGGCTTCCCAAGATGTTATCAAGGAGTTTAAAGGTTTAAAGGTTTCGATCGAAGGA 60
Db 1 ATGAGGCTTCCCAAGATGTTATCAAGGAGTTTAAAGGTTTAAAGGTTTCGATCGAAGGA 60
QY 61 ACGGTCAATGGCCACGAGTTTGAATAGAGGCGAAGAGAGAGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGCCACGAGTTTGAATAGAGGCGAAGAGAGAGGAGGCCATACGAAGGC 120
QY 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGACCCCTGCCGACATACCA 240
Db 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGACCCCTGCCGACATACCA 240
QY 241 GACTATATAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATATAAGCTGTCTATTTCTTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300

Qy 301 GACGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGGTCAGGTTCAATGGCGTGAACCTTTCCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCAGGTTCAATGGCGTGAACCTTTCCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 421 ATGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTGATGGCGTTTGAAGAGAGAG 480
Db 421 ATGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTGATGGCGTTTGAAGAGAGAG 480
Qy 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Db 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Qy 541 TACATGGCAAGAAGCCCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAGAAGCCCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATAAACGCCCAACGAAGACTATACAACTGTTGAGCAGTATGAAGAACCGAGGAGCG 660
Db 601 ATAAACGCCCAACGAAGACTATACAACTGTTGAGCAGTATGAAGAACCGAGGAGCG 660
Qy 661 CACCATCTGTTCCCTTTAA 678
Db 661 CACCATCTGTTCCCTTTAA 678

RESULT 7

US-10-311-030-5

; Sequence 5, Application US/10311030

; Publication No. US20040171107A1

; GENERAL INFORMATION:

; APPLICANT: Nelson, David

; APPLICANT: Zamiatra, Elize

; APPLICANT: Tsien, Roger

; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS

; FILE REFERENCE: 15916-032U51

; CURRENT APPLICATION NUMBER: US/10/311,030

; CURRENT FILING DATE: 2002-12-10

; PRIOR APPLICATION NUMBER: PCT/US01/04625

; PRIOR FILING DATE: 2001-02-13

; PRIOR APPLICATION NUMBER: US 60/184,732

; PRIOR FILING DATE: 2000-02-23

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 5

; LENGTH: 678

; TYPE: DNA

; ORGANISM: Discosoma sp.

US-10-311-030-5

Query Match 100.0%; Score 678; DB 19; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216; Mismatches 0; Indels 0; Gaps 0;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
Qy 61 ACGGTCAATGGGACGAGTTTGAATAGAGCGCAAGGAGAGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGGACGAGTTTGAATAGAGCGCAAGGAGAGGGAGGCCATACGAAGGC 120
Qy 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 240
Db 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 240

Qy 241 GACTATAAAAGCTGTCTATTTCTGAGGATTTAAATGGGAAGGGTCATGAACTTTGAA 300
Db 241 GACTATAAAAGCTGTCTATTTCTGAGGATTTAAATGGGAAGGGTCATGAACTTTGAA 300
Qy 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Qy 361 AAGTCAAGTTCATTTGCGTGAACCTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGTCAAGTTCATTTGCGTGAACCTTTCTTCCGATGGACCTGTTATGCAAAAGAGACA 420
Qy 421 ATGGGCTGGGAAGCCAGCAGCTGAGCGTTTCTATCCTCGTATGGCGTTTGAAGAGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCAGCTGAGCGTTTCTATCCTCGTATGGCGTTTGAAGAGAGAG 480
Qy 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Db 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCAATACCTAGTTGAATTCAAAAGTATT 540
Qy 541 TACATGGCAAGAAGCCCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGGCAAGAAGCCCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT 600
Qy 601 ATAAACGCCCAACGAAGACTATACAACTGTTGAGCAGTATGAAGAACCGAGGAGCG 660
Db 601 ATAAACGCCCAACGAAGACTATACAACTGTTGAGCAGTATGAAGAACCGAGGAGCG 660
Qy 661 CACCATCTGTTCCCTTTAA 678
Db 661 CACCATCTGTTCCCTTTAA 678

RESULT 8

US-10-656-029-21

; Sequence 21, Application US/10656029

; Publication No. US20050003367A1

; GENERAL INFORMATION:

; APPLICANT: VERTEX PHARMACEUTICALS INC.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RAPID DEVELOPMENT OF

; FILE REFERENCE: VPI/02-143WO2

; CURRENT APPLICATION NUMBER: US/10/656,029

; CURRENT FILING DATE: 2003-09-05

; PRIOR APPLICATION NUMBER: 60/408,297

; PRIOR FILING DATE: 2002-09-05

; NUMBER OF SEQ ID NOS: 86

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 21

; LENGTH: 678

; TYPE: DNA

; ORGANISM: Discosoma sp.

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(678)

; OTHER INFORMATION: fluorescent protein

US-10-656-029-21

Query Match 100.0%; Score 678; DB 21; Length 678;
Best Local Similarity 100.0%; Pred. No. 5.9e-216; Mismatches 0; Indels 0; Gaps 0;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
Qy 61 ACGGTCAATGGGACGAGTTTGAATAGAGCGCAAGGAGAGGGAGGCCATACGAAGGC 120
Db 61 ACGGTCAATGGGACGAGTTTGAATAGAGCGCAAGGAGAGGGAGGCCATACGAAGGC 120
Qy 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 121 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATT 180

```
QY 181 TTGTACCAACAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCGACATACCA 240
Db 181 TTGTACCAACAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCGACATACCA 240
QY 241 GACTATAAAAGCTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 241 GACTATAAAAGCTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
QY 301 GACGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 301 GACGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
QY 361 AAGGTCAAGTTCATTGCGGTGAACCTTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 361 AAGGTCAAGTTCATTGCGGTGAACCTTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
QY 421 ATGGGCTGGGAAGCCAGCACTAGCGTTTGTATCCTCGTATGCGGTGTTGAAAGGAGAG 480
Db 421 ATGGGCTGGGAAGCCAGCACTAGCGTTTGTATCCTCGTATGCGGTGTTGAAAGGAGAG 480
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 540
Db 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 540
QY 541 TACATGCAAGAGAGCGTGTGACGTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 541 TACATGCAAGAGAGCGTGTGACGTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
QY 601 ATAAAGGCCACACGAAAGACTATACAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
Db 601 ATAAAGGCCACACGAAAGACTATACAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
QY 661 CACCATCTGTTCTTTAA 678
Db 661 CACCATCTGTTCTTTAA 678

RESULT 9
US-09-999-745-66
; Sequence 66, Application US/09999745
; Patent No. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; PRIOR FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 66
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-999-745-66

Query Match 100.0%; Score 678; DB 9; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
Db 54 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGGTTTAAGGTTTCGATCGAAGGA 113
QY 61 ACGTCAATGGCGACGAGTTTGAATAGAGCGAAGGAGGAGGCCATACGAAGGC 120
```

```
Db 114 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGGAGAGGGAGGCCATACGAAGGC 173
QY 121 CACAATACCGTAAAGCTTAAAGGTAAACCAAGGGGGACCTTTGGCCATTTGCTTGGGATATT 180
Db 174 CACAATACCGTAAAGCTTAAAGGTAAACCAAGGGGGACCTTTGGCCATTTGCTTGGGATATT 233
QY 181 TTGTACCAACAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCGACATACCA 240
Db 234 TTGTACCAACAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCGACATACCA 293
QY 241 GACTATAAAAGCTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 294 GACTATAAAAGCTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 353
QY 301 GACGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 354 GACGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 413
QY 361 AAGGTCAAGTTCATTGCGGTGAACCTTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 414 AAGGTCAAGTTCATTGCGGTGAACCTTTCCTCCGATGGACCTGTTATGCAAAAGAGACA 473
QY 421 ATGGGCTGGGAAGCCAGCACTAGCGTTTGTATCCTCGTATGCGGTGTTGAAAGGAGAG 480
Db 474 ATGGGCTGGGAAGCCAGCACTAGCGTTTGTATCCTCGTATGCGGTGTTGAAAGGAGAG 533
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 540
Db 534 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGGTCAATTACCTAGTTGAATTCAAAAGTATT 593
QY 541 TACATGCAAGAGAGCGTGTGACGTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 594 TACATGCAAGAGAGCGTGTGACGTACCAAGGCTACTATGTTGACTCCAAACTGGAT 653
QY 601 ATAAAGGCCACACGAAAGACTATACAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 660
Db 654 ATAAAGGCCACACGAAAGACTATACAATCGTTGAGCAGTATGAAGAACCGAGGGACGC 713
QY 661 CACCATCTGTTCTTTAA 678
Db 714 CACCATCTGTTCTTTAA 731

RESULT 10
US-09-866-538-11
; Sequence 11, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-866-538-11

Query Match 100.0%; Score 678; DB 10; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
Db 54 ATGAGGCTTCCAAAGATGTTATCAAGGATTCATGAGGTTTAAGGTTTCGATCGAAGGA 113
```

QY 61 ACGGTCAATGGGACGAGTTTGAATAGAGCGGAAGGAGGAGGGGAGGCCATACGAAGGC 120
DB 114 ACGGTCAATGGGACGAGTTTGAATAGAGCGGAAGGAGGAGGGGAGGCCATACGAAGGC 173
QY 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTGCTTGGGATATT 180
DB 174 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGACCTTTGCCATTGCTTGGGATATT 233
QY 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 240
DB 234 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 293
QY 241 GACTATAAAAGCTGTCATTCTTCAAGAGGATTTAAATGGGAAGGGTCAATGAACCTTTGAA 300
DB 294 GACTATAAAAGCTGTCATTCTTCAAGAGGATTTAAATGGGAAGGGTCAATGAACCTTTGAA 353
QY 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTATGGAAGGAGGAGGAGGAG 360
DB 354 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTATGGAAGGAGGAGGAGGAG 413
QY 361 AAGTCAAGTTCATTGGCGTGAACCTTTCCATGGAACCTGTTATGCAAGAGAGCA 420
DB 414 AAGTCAAGTTCATTGGCGTGAACCTTTCCATGGAACCTGTTATGCAAGAGAGCA 473
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGCGGTTGAAAGGAGAG 480
DB 474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGCGGTTGAAAGGAGAG 533
QY 481 ATTCAATAGGCTCGAAGCTGAAGAGCGGTGCTACCTACCTAGTTGAATTCMAAGTATT 540
DB 534 ATTCAATAGGCTCGAAGCTGAAGAGCGGTGCTACCTACCTAGTTGAATTCMAAGTATT 593
QY 541 TACATGCAAGAGCGCTGTCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
DB 594 TACATGCAAGAGCGCTGTCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 653
QY 601 ATAAAGGCCACAAAGAGCTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGAGCG 660
DB 654 ATAAAGGCCACAAAGAGCTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGAGCG 713
QY 661 CACCATCTGTTCTCTTTAA 678
DB 714 CACCATCTGTTCTCTTTAA 731

RESULT 11

US-09-794-308-11
; Sequence 11, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: TSIEN, Roger
; APPLICANT: ZACHARIAS, David
; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-794-308-11

Query Match 100.0%; Score 678; DB 10; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGGTCTTCCAAAGAAATGTTTCAAGGAGTTCATGAGGTTTAAAGGTTCCGATCGGAAGGA 60
DB 54 ATGAGGTCTTCCAAAGAAATGTTTCAAGGAGTTCATGAGGTTTAAAGGTTCCGATCGGAAGGA 113
QY 61 ACGGTCAATGGGACGAGTTTGAATAGAGCGGAAGGAGGAGGGGAGGCCATACGAAGGC 120
DB 114 ACGGTCAATGGGACGAGTTTGAATAGAGCGGAAGGAGGAGGGGAGGCCATACGAAGGC 173
QY 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGAGCCTTTGCCATTGCTTGGGATATT 180
DB 174 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGAGCCTTTGCCATTGCTTGGGATATT 233
QY 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 240
DB 234 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCGAAGCAACCTGCCGACATACCA 293
QY 241 GACTATAAAAGCTGTCATTCTTCAAGAGGATTTAAATGGGAAGGGTCAATGAACCTTTGAA 300
DB 294 GACTATAAAAGCTGTCATTCTTCAAGAGGATTTAAATGGGAAGGGTCAATGAACCTTTGAA 353
QY 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTATGGAAGGAGGAGGAGGAG 360
DB 354 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTATGGAAGGAGGAGGAGGAG 413
QY 361 AAGTCAAGTTCATTGGCGTGAACCTTTCCATGGAACCTGTTATGCAAGAGAGCA 420
DB 414 AAGTCAAGTTCATTGGCGTGAACCTTTCCATGGAACCTGTTATGCAAGAGAGCA 473
QY 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGCGGTTGAAAGGAGAG 480
DB 474 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGCGGTTGAAAGGAGAG 533
QY 481 ATTCAATAGGCTCGAAGCTGAAGAGCGGTGCTACCTACCTAGTTGAATTCMAAGTATT 540
DB 534 ATTCAATAGGCTCGAAGCTGAAGAGCGGTGCTACCTAGTTGAATTCMAAGTATT 593
QY 541 TACATGCAAGAGCGCTGTCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
DB 594 TACATGCAAGAGCGCTGTCAGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 653
QY 601 ATAAAGGCCACAAAGAGCTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGAGCG 660
DB 654 ATAAAGGCCACAAAGAGCTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGAGCG 713
QY 661 CACCATCTGTTCTCTTTAA 678
DB 714 CACCATCTGTTCTCTTTAA 731

RESULT 12

US-09-865-291-11
; Sequence 11, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
; APPLICANT: TSIEN, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-09-865-291-11

```
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
; US-10-433-640-12

Query Match      100.0%; Score 678; DB 10; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGCTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 60
Db 54 ATGAGGCTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 113
Qy 61 ACGGTCATAGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 120
Db 114 ACGGTCATAGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 173
Qy 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGCCATTTTGGTTCGGATATT 180
Db 174 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGCCATTTTGGTTCGGATATT 233
Qy 181 TTGTCACCAAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCAACCTTGCCTGCGACATACCA 240
Db 234 TTGTCACCAAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCAACCTTGCCTGCGACATACCA 293
Qy 241 GACTATAAAAAGCTGTCATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 294 GACTATAAAAAGCTGTCATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 353
Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
Db 354 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 413
Qy 361 AAGGTCAAAGTTCATTGCGGTGAACCTTTCCCTCCGATGACCTGTTATGCAAAAAGAGACA 420
Db 414 AAGGTCAAAGTTCATTGCGGTGAACCTTTCCCTCCGATGACCTGTTATGCAAAAAGAGACA 473
Qy 421 ATGGGCTGGGAGCCAGCAGCTGTTGATCCTCGTGAATGCGCTGTTGAAAGGAGAG 480
Db 474 ATGGGCTGGGAGCCAGCAGCTGTTGATCCTCGTGAATGCGCTGTTGAAAGGAGAG 533
Qy 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCAATTCCTAGTTGAAATTCAAAAGTATT 540
Db 534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCAATTCCTAGTTGAAATTCAAAAGTATT 593
Qy 541 TACATGCAAAAGAGCTGTGACGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 594 TACATGCAAAAGAGCTGTGACGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 653
Qy 601 ATACAGCCACACGAGACTATACAACTCGTTGAGCAGTATGAAGAACCGAGGACGC 660
Db 654 ATACAGCCACACGAGACTATACAACTCGTTGAGCAGTATGAAGAACCGAGGACGC 713
Qy 661 CACCATCTGTTCCCTTTAA 678
Db 714 CACCATCTGTTCCCTTTAA 731

RESULT 13
US-10-433-640-12
; Sequence 12, Application US/10433640
; Publication No. US20040115792A1
; GENERAL INFORMATION:
; APPLICANT: Lichtenberg-Frate, Hella
; TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
; FILE REFERENCE: 1487/3
; CURRENT APPLICATION NUMBER: US/10/433,640
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/14610
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: DE 10061872.3
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 859
```

```
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
; US-10-433-640-12

Query Match      100.0%; Score 678; DB 19; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGCTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 60
Db 54 ATGAGGCTCTTCCAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 113
Qy 61 ACGGTCATAGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 120
Db 114 ACGGTCATAGGGACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGC 173
Qy 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGCCATTTTGGTTCGGATATT 180
Db 174 CACATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGCCATTTTGGTTCGGATATT 233
Qy 181 TTGTCACCAAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCAACCTTGCCTGCGACATACCA 240
Db 234 TTGTCACCAAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCAACCTTGCCTGCGACATACCA 293
Qy 241 GACTATAAAAAGCTGTCATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 294 GACTATAAAAAGCTGTCATTTCTTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 353
Qy 301 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 360
Db 354 GACGGTGGCGTCTGTTACTGTAACCCAGGATTTCCAGTTTTCAGGATGGCTGTTTCATCTAC 413
Qy 361 AAGGTCAAAGTTCATTGCGGTGAACCTTTCCCTCCGATGACCTGTTATGCAAAAAGAGACA 420
Db 414 AAGGTCAAAGTTCATTGCGGTGAACCTTTCCCTCCGATGACCTGTTATGCAAAAAGAGACA 473
Qy 421 ATGGGCTGGGAGCCAGCAGCTGTTGATCCTCGTGAATGCGCTGTTGAAAGGAGAG 480
Db 474 ATGGGCTGGGAGCCAGCAGCTGTTGATCCTCGTGAATGCGCTGTTGAAAGGAGAG 533
Qy 481 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCAATTCCTAGTTGAAATTCAAAAGTATT 540
Db 534 ATTCATAAGGCTCTGAAGCTGAAAGACGGTGGTCAATTCCTAGTTGAAATTCAAAAGTATT 593
Qy 541 TACATGCAAAAGAGCTGTGACGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 600
Db 594 TACATGCAAAAGAGCTGTGACGCTACCAAGGCTACTATGTTGACTCCAAACTGGAT 653
Qy 601 ATACAGCCACACGAGACTATACAACTCGTTGAGCAGTATGAAGAACCGAGGACGC 660
Db 654 ATACAGCCACACGAGACTATACAACTCGTTGAGCAGTATGAAGAACCGAGGACGC 713
Qy 661 CACCATCTGTTCCCTTTAA 678
Db 714 CACCATCTGTTCCCTTTAA 731

RESULT 14
US-10-885-988-11
; Sequence 11, Application US/10885988
; Publication No. US20040259165A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/10/885,988
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/866,538
```

; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-10-885-988-11

Query Match 100.0%; Score 678; DB 20; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216; Indels 0; Gaps 0;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGCTCCCAAGAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
Db |||||||
Qy 54 ATGAGGCTCCCAAGAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 113
Db |||||||
Qy 61 ACGGTCAATGGGCGACGAGTTTGAATAGAGCGGAAGAGAGGGAGGCCATACGAAGGC 120
Db |||||||
Qy 114 ACGGTCAATGGGCGACGAGTTTGAATAGAGCGGAAGAGAGGGAGGCCATACGAAGGC 173
Db |||||||
Qy 121 CACATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db |||||||
Qy 174 CACATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 233
Db |||||||
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGACCCCTGCCGACATACCA 240
Db |||||||
Qy 234 TTGTCCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGACCCCTGCCGACATACCA 293
Db |||||||
Qy 241 GACTATATAAAGCTGTCAATTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300
Db |||||||
Qy 294 GACTATATAAAGCTGTCAATTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353
Db |||||||
Qy 301 GACGCTGGCGCTGCTTACTGTAACCCAGGATTCAGATTTGCAAGATGGCTGTTTCATCTAC 360
Db |||||||
Qy 354 GACGCTGGCGCTGCTTACTGTAACCCAGGATTCAGATTTGCAAGATGGCTGTTTCATCTAC 413
Db |||||||
Qy 361 AAGGTCAAGTTCATTTGGGCTGAACTTTCTTCCGATGACCTGTTATGCAAAAGAGACA 420
Db |||||||
Qy 414 AAGGTCAAGTTCATTTGGGCTGAACTTTCTTCCGATGACCTGTTATGCAAAAGAGACA 473
Db |||||||
Qy 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCCTCGTATGGCGTGTGTAAGAGGAGAG 480
Db |||||||
Qy 474 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCCTCGTATGGCGTGTGTAAGAGGAGAG 533
Db |||||||
Qy 481 ATTCATAAGGCTCTGAAGCTGAAAGAGCGGTGGTCAATTCCTAGTGTGAATTCAAAAGTATT 540
Db |||||||
Qy 534 ATTCATAAGGCTCTGAAGCTGAAAGAGCGGTGGTCAATTCCTAGTGTGAATTCAAAAGTATT 593
Db |||||||
Qy 541 TACATGGCAAGAAGCCCTGTCAGCTACAGGGTACTATGTTGATCTCCAAACTCGAT 600
Db |||||||
Qy 594 TACATGGCAAGAAGCCCTGTCAGCTACAGGGTACTATGTTGATCTCCAAACTCGAT 653
Db |||||||
Qy 601 ATAAACAGCCCAACAGAGACTATACATTCGTTGAGCAGTATGAAAGAACCGAGGGAGCG 660
Db |||||||
Qy 654 ATAAACAGCCCAACAGAGACTATACATTCGTTGAGCAGTATGAAAGAACCGAGGGAGCG 713
Db |||||||
Qy 661 CACCATCTGTTCTTTAA 678
Db |||||||
Qy 714 CACCATCTGTTCTTTAA 731
Db |||||||

RESULT 15
US-10-857-622-11
; Sequence 11, Application US/10857622
; Publication No. US20050026234A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: VIOLIN, Jonathan
; APPLICANT: NEWTON, Alexandra

; APPLICANT: TSIEH, Roger
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: Emission Ratiometric Indicators of Phosphorylation By C-Kinase
; FILE REFERENCE: 39754-0891 CPCICP2
; CURRENT APPLICATION NUMBER: US/10/857,622
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: US 09/865,291
; PRIOR FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: US 09/396,003
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: US 08/792,553
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: US 594,575
; PRIOR FILING DATE: 1996-01-31
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 859
; TYPE: DNA
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (54)..(731)
US-10-857-622-11

Query Match 100.0%; Score 678; DB 21; Length 859;
Best Local Similarity 100.0%; Pred. No. 6.7e-216; Indels 0; Gaps 0;
Matches 678; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAGGCTCTCCAGGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
Db |||||||
Qy 54 ATGAGGCTCTCCAGGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 113
Db |||||||
Qy 61 ACGGTCAATGGGCGACGAGTTTGAATAGAGCGGAAGAGAGGGAGGCCATACGAAGGC 120
Db |||||||
Qy 114 ACGGTCAATGGGCGACGAGTTTGAATAGAGCGGAAGAGAGGGAGGCCATACGAAGGC 173
Db |||||||
Qy 121 CACATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db |||||||
Qy 174 CACATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 233
Db |||||||
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGACCCCTGCCGACATACCA 240
Db |||||||
Qy 234 TTGTCCACCAATTTTCAGTATGGAAGCAAGGATATATGTCAGACCCCTGCCGACATACCA 293
Db |||||||
Qy 241 GACTATATAAAGCTGTCAATTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 300
Db |||||||
Qy 294 GACTATATAAAGCTGTCAATTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAACTTTGAA 353
Db |||||||
Qy 301 GACGCTGGCGCTGCTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db |||||||
Qy 354 GACGCTGGCGCTGCTTACTGTAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 413
Db |||||||
Qy 361 AAGGTCAAGTTCATTTGGGCTGAACTTTCTTCCGATGACCTGTTATGCAAAAGAGACA 420
Db |||||||
Qy 414 AAGGTCAAGTTCATTTGGGCTGAACTTTCTTCCGATGACCTGTTATGCAAAAGAGACA 473
Db |||||||
Qy 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCCTCGTATGGCGTGTGTAAGAGGAGAG 480
Db |||||||
Qy 474 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCCTCGTATGGCGTGTGTAAGAGGAGAG 533
Db |||||||
Qy 481 ATTCATAAGGCTCTGAAGCTGAAAGAGCGGTGGTCAATTCCTAGTGTGAATTCAAAAGTATT 540
Db |||||||
Qy 534 ATTCATAAGGCTCTGAAGCTGAAAGAGCGGTGGTCAATTCCTAGTGTGAATTCAAAAGTATT 593
Db |||||||
Qy 541 TACATGGCAAGAAGCCCTGTCAGCTACAGGGTACTATGTTGATCTCCAAACTCGAT 600
Db |||||||
Qy 594 TACATGGCAAGAAGCCCTGTCAGCTACAGGGTACTATGTTGATCTCCAAACTCGAT 653
Db |||||||
Qy 601 ATAAACAGCCCAACAGAGACTATACATTCGTTGAGCAGTATGAAAGAACCGAGGGAGCG 660
Db |||||||
Qy 654 ATAAACAGCCCAACAGAGACTATACATTCGTTGAGCAGTATGAAAGAACCGAGGGAGCG 713
Db |||||||

Qy	661	CACCATCTGTTCCCTTAA	678
Db	714	CACCATCTGTTCCCTTAA	731

Search completed: July 1, 2005, 18:32:40
Job time : 573 secs

Result No.	Score	Match	%	Length	DB	ID	Description	
1	678	100.0		678	3	US-09-459-956-6	Sequence 6, Appli	
2	450.6	66.5		723	4	US-10-152-296-1	Sequence 1, Appli	
3	409.6	60.4		699	3	US-09-459-956-5	Sequence 5, Appli	
4	271	40.0		801	3	US-09-459-956-7	Sequence 7, Appli	
5	178.4	26.3		690	3	US-09-459-956-2	Sequence 2, Appli	
6	147.2	21.7		696	3	US-09-459-956-4	Sequence 4, Appli	
7	141.4	20.9		1482	4	US-09-977-897-1	Sequence 1, Appli	
8	133.4	19.7		696	3	US-09-459-956-3	Sequence 3, Appli	
9	133.2	19.6		1079	3	US-09-609-161B-15	Sequence 15, Appl	
10	133.2	19.6		1079	4	US-09-626-581D-64	Sequence 64, Appl	
11	133.2	19.6		1079	4	US-09-415-765B-84	Sequence 84, Appl	
12	133.2	19.6		1079	4	US-09-626-580C-64	Sequence 64, Appl	
13	133.2	19.6		1085	3	US-09-277-716-15	Sequence 15, Appl	
14	132.4	19.5		1104	3	US-09-277-716-30	Sequence 30, Appl	
15	132.4	19.5		1104	3	US-09-609-161B-30	Sequence 30, Appl	
16	132.4	19.5		1279	3	US-09-277-716-31	Sequence 31, Appl	
17	132.4	19.5		1279	3	US-09-609-161B-31	Sequence 31, Appl	
18	130	19.2		1021	4	US-09-839-650-2	Sequence 2, Appli	
19	128.4	18.9		720	4	US-09-839-650-1	Sequence 1, Appli	
20	42	6.2		7218	1	US-08-232-463-14	Sequence 14, Appl	
21	35.8	5.3		1141	4	US-09-806-708B-22	Sequence 22, Appl	
22	34.4	5.1		2748	4	US-09-540-236-1106	Sequence 1106, Ap	
23	34.4	5.1		19888	4	US-09-596-002-10	Sequence 10, Appl	
24	34.2	5.0		14207	4	US-09-949-016-12187	Sequence 12187, A	
25	34.2	5.0		14207	4	US-09-949-016-17460	Sequence 17460, A	
26	33.6	5.0		591	4	US-09-248-796A-2585	Sequence 2585, Ap	
27	33.2	4.9		601	4	US-09-949-016-43022	Sequence 43022, A	

301	Qy	GACGGTGGCGTCGTTACTGTGTAAACCCAGGATTCACGTTTCAGATGGCTGTTTCATCTAC	360
301	Db	GACGGTGGCGTCGTTACTGTGTAAACCCAGGATTCACGTTTCAGATGGCTGTTTCATCTAC	360
361	Qy	AAGGTCAAGTTTCATTGGCGTGAACTTCCTCCGATGGACCTGTTATGCAAAAGAGACA	420
361	Db	AAGGTCAAGTTTCATTGGCGTGAACTTCCTCCGATGGACCTGTTATGCAAAAGAGACA	420
421	Qy	ATGGGCTGGGAAGCCAGCACTGACGCTTTGTATCCTCGTGATGGCGTGTGAAAGGAGAG	480
421	Db	ATGGGCTGGGAAGCCAGCACTGACGCTTTGTATCCTCGTGATGGCGTGTGAAAGGAGAG	480
481	Qy	ATTCATAAAGGCTCTGAAGCTGAAAGACGGTGCTATTACCTAGTTGTAATTCAAAAGTATT	540
481	Db	ATTCATAAAGGCTCTGAAGCTGAAAGACGGTGCTATTACCTAGTTGTAATTCAAAAGTATT	540
541	Qy	TACATGGCAAGACGCTGTCAGCTACACGGTACTACTATGTTGACTCCAAACTGGAT	600
541	Db	TACATGGCAAGACGCTGTCAGCTACACGGTACTACTATGTTGACTCCAAACTGGAT	600
601	Qy	ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGACGC	660
601	Db	ATAACAAGCCACAACGAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGACGC	660
661	Qy	CACCATCTGTTCCCTTTAA	678
661	Db	CACCATCTGTTCCCTTTAA	678

```

RESULT 2
US-10-152-296-1
Sequence 1, Application US/10152296
Patent No. 6723537
GENERAL INFORMATION:
APPLICANT: Peelle, Beau
APPLICANT: Rigel Pharmaceuticals, Incorporated
TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells
FILE REFERENCE: 021044-000110US
CURRENT APPLICATION NUMBER: US/10/152,296
CURRENT FILING DATE: 2002-12-10
PRIOR APPLICATION NUMBER: US 60/291,871
PRIOR FILING DATE: 2001-05-18
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 723
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:mammalian
OTHER INFORMATION: codon-optimized variant (DsRED) of Discosoma sp.
OTHER INFORMATION: "red" red fluorescent protein (RFP)
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(723)
OTHER INFORMATION: DsRED
US-10-152-296-1

```

[illegible]

```

RESULT 3
US-09-459-956-5
; Sequence 5, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, Ili, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; TITLE OF INVENTION: OPTICAL METHODS
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 699
; TYPE: DNA
; ORGANISM: Discosoma striata
US-09-459-956-5

```

Query Match	60.4%	Score	409.6	DB	3	Length	699
Best Local Similarity	76.2%	Pred. No.	66-134				
Matches	518	Conservative	0	Mismatches	159	Indels	3
Gaps							

Qy	1	ATGAGGTTCTCCAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTTCGATCGAAGGA	60
Db	1	ATGAGTTGTGTTCCAGAGTGTGATCAAGGAAGAAATGTTGATCGATCTTCATCTCTCGAAGGA	60
Oy	61	ACGGTCAATGGGCACGAGTTTGAATAGAAAGGCCAAGGAGAGGGGAGGCCATACGAAGGC	120


```
; ORGANISM: Anemonia majano
US-09-459-956-2

Query Match      26.3%; Score 178.4; DB 3; Length 690;
Best Local Similarity 56.1%; Pred. No. 3.3e-52;
Matches 360; Conservative 0; Mismatches 276; Indels 6; Gaps 1;

Qy 9 TTCAAGAATGTTATCAAGGAGTTCATGAGGTTTAAAGTTCGCTGGAAGAACGGTCAA 68
Db 9 TTCAACAAGTTTATCGGAGATGACATGAAATGACCTACCATATGATGCTGTGCAA 68
Qy 69 TGGCAGCAGGTTTGAATAGAGCGAAGAGAGAGGGAGGCCATACGAAGGCCACAATAC 128
Db 69 TGGCAGTTACTTTACCGTCAAAGTGAAGGCAACGGGAAGCCATACGAAGGAGCGCAGAC 128
Qy 129 CGTAAGCTTAAAGTAA-----CCAGGGGGGACCTTTGCGCATTTGCTGGGATATTTT 182
Db 129 TTCGACTTTTAAAGTCACCATGCGCCACCGTGGGCCCCCTTGCAATCTCTTTGACATAC 188
Qy 183 GTCACCAAAATTCAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCAGA 242
Db 189 ATCTACAGTGTTCAAATATGAAATCGATGCTTTACTGCGTATCTTACCAGTATGCCCGA 248
Qy 243 CTATAAAAGCTGTCTATTCCTGGAAGGATTTAAATGGGAAGGTCATGAACCTTTGAAGA 302
Db 249 CTATTTCAACAAGCATTTCTGACGGAAATGCTCATATGAAAGGACTTTTACCTATGAAGA 308
Qy 303 CGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGCAGATGCGTGTTCATCTACAA 362
Db 309 TGSAGGAGTTGCTACAGCCAGTTGGGAAATTAAGCCCTTAAAGGCCAAGTCTTTGAGCACAA 368
Qy 363 GGTCAAGTTTCAATGGCGTGAACCTTTCTCCGATGGACCTGTATGCAAAAGAGACAAAT 422
Db 369 ATCCACGTTTCATGGAGTGAACCTTTCTCGTATGGACCTGTATGGGAAGAGAACAC 428
Qy 423 GGGCTGGGAAGCCAGCACTAGCGCTTTGTATCTCTCGTATGCGGTGTGTGAAGAGAGAT 482
Db 429 TGGTTGGACCCATCTTTTGAGAAATGACTGTCTGCGATGGAATATTGAAGGGTGTATG 488
Qy 483 TCATAAGCTCTGAAGCTGAAGAGACGCTGTGTCATTACTAGTTGAATTCAAAAGTATTTA 542
Db 489 CACCGCTTCTCATGCTGCAAGGAGGTGGCAATTTACAGATGGCAATTTCCACACTTCTTA 548
Qy 543 CATGGCAAAAGCCTGTGACGTACACAGGGTACTTACTATGTGACTCCAAACTGGGATAT 602
Db 549 CAAGACAAAAAACCGGTGACGTGCCACCAACCATGTGTGTGGAAACATCCATTTGGAG 608
Qy 603 AACAGGCCACAAACGAAGACTATACAATCGTTGAGCAGTATGA 644
Db 609 GACCGACCTTGACAAAGGTGGCAACAGTGTTCAGCTGACGGA 650
```

```
RESULT 6
US-09-459-956-4
; Sequence 4, Application US/09459956
; Patent No. 6342379
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Gonzalez, III, Jesus E.
; TITLE OF INVENTION: DETECTION OF TRANSMEMBRANE POTENTIALS BY
; FILE REFERENCE: REGEN1290-4
; CURRENT APPLICATION NUMBER: US/09/459,956
; CURRENT FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: 08/765,860
; PRIOR FILING DATE: 1999-05-08
; PRIOR APPLICATION NUMBER: 08/481,977
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: PCT/US96/09652
; PRIOR FILING DATE: 1996-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
```

```
; LENGTH: 696
; TYPE: DNA
; ORGANISM: Zoanthus sp
US-09-459-956-4

Query Match      21.7%; Score 147.2; DB 3; Length 696;
Best Local Similarity 56.8%; Pred. No. 3.5e-41;
Matches 316; Conservative 0; Mismatches 228; Indels 12; Gaps 2;

Qy 9 TTCCAGAATGTTATCAAGGAGTTCATGAGGTTTAAAGTTCGCTGGAAGAACGGTCAA 68
Db 9 TTCAAGACACGGTCTAAAGAGAAATGACAAATACCAATGAAATGGAAGGGTGCSTCAA 68
Qy 69 TGGCAGCAGGTTTGAATAGAGCGAAGAGAGAGGGAGGCCATACGAAGGCCACAATAC 128
Db 69 CGACACATAAATTTGTGATCAGGGCGAGGCGATTTGGATATCCGTTCAAGGGGAACAGAC 128
Qy 129 CGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTTGCCATTTGCTGGGATATTTTGTCAAC 188
Db 129 TATTAATCTGTGTGATCGAAGGGGGACCATTTGCCATTTTCCGAAGACATATTTGTGAGC 188
Qy 189 ACAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCAGACTATA 248
Db 189 TGGCTTTTAAGTACGCGAGACAGGATTTTCACTGAATATCTCTCAAGACATAGTAGACTAT 248
Qy 249 AAAGCTGTCTATTTCTCTGAAGGATTTAAATGGGAAAGGGTCAATGAACTTTGAAGACGGTG 307
Db 249 CAAGAAGCTCGTGTCTGCTGGATATACATGCGGCGAGTCTTTTCTCTTTGAGATGGAGC 308
Qy 308 -----CGCTCGTTACTGTAAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTCATCTACAA 362
Db 309 AGTCTGCATATGCAATGTAGATATAACAGTGTGTCAAGAGAACTGCAATTTATCATAA 368
Qy 363 GGTCAAGTTTCAATGGCGTGAACCTTTCTCCGATGACCTGTTCATGCAAAAGAGACAAAT 422
Db 369 GAGCATATTTAATGGAATGAAATTTCTCTGCTGATGAGACCTGTGATGAAAGAGATGACAC 428
Qy 423 GGGCTGGGAAGCCAGCACTAGCGCTTT-----GTATCTCTGCTGATGGCGGTGTGAAAGG 476
Db 429 TAACTGGGAAGCATCTCGCGAAGATCATGCCAGTACCTTAAGCAGGGGATCTGAAAGG 488
Qy 477 AGAGATTCATAGGCTGTGAAGCTGAAAGACGGTGTGTCATTACTTAGTTGAAATTCAAAAG 536
Db 489 GGATGCTCCATGACTCTCTCTCTGAAGGATGTGGGCGTTACCGGTTGCCAGTTCGACAC 548
Qy 537 TATTACATGGCCAAAG 552
Db 549 AGTTTACAAAGCAAAG 564
```

```
RESULT 7
US-09-977-897-1
; Sequence 1, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1482
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-1
```

```
Query Match      20.9%; Score 141.4; DB 4; Length 1482;
Best Local Similarity 51.3%; Pred. No. 6.5e-39;
Matches 328; Conservative 0; Mismatches 311; Indels 0; Gaps 0;
```



```
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-15

Query Match      19.6%; Score 133.2; DB 3; Length 1079;
Best Local Similarity 50.6%; Pred. No. 4.2e-36;
Matches 321; Conservative 0; Mismatches 313; Indels 0; Gaps 0;

Qy 13 AAGATGTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGAACGGTCAATGGG 72
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 280 AACACTTGTGTTACAAAGTAATGTCGTATAAGTAATCTCGAAGGAATTTGTAACAC 339
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 73 CACGAGTTTGAATAGAACGCGAAGGAGGAGGCCATACGAAGGCCAACATACCGTA 132
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 340 CATGTTTTCACAAAGAGTAATGTCGTATAAGTAATCTCGAAGGAATTTGTAACAC 399
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 133 AAGCTTAAAGTAAACAAAGGGGGACCTTTGCCATTTGGTGGGATATTTTGTCAACCAA 192
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 400 CAGATTCTGTGCACGAAGGGGCCCACTGCTTTTGCATTTGATATTTGTGCACCACT 459
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 193 TTTTCAGTATGAACAGGATATATGTCAGCACCTGCCGACATACCAAGCTATAAAAG 252
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 460 TTTCAATATGCAACCGTACTTTTCACGAATATCCGAATGATATATCAGATTATTTATA 519
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 253 CTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTTCATGAACCTTTGAAGACGGTGGCGTC 312
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 520 CAATCATTTCCAGCAGAGTTTATGATGAACGAACATTCAGTTACGAAGATGGCGACTT 579
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 313 GTTACTGTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTCAAGGTCAGGTC 372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 580 GTTCAATTCGTTCCAGAGATGATGATGAACGAACATTCAGTTACGAAGATGGCGACTT 639
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 373 ATTGGCGTGAACCTTCCCTCGATGGCGTGTGAAGGAGAGATTCATAAGGCT 492
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 700 CCTTCATTGAAGCCATGATCATGATGAATGAGCTGTTGGTCGGGAAGTAATCTTGTG 759
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 493 CTGAAGCTGAAGACGGTGGTTCATTAAGTGAATTCAGGATTTTCAATGCGCAAG 552
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 760 TATAAATTAACCTCTGGGAATATTAATCATGTACATGAAACATTAATGAAGTCGAA 819
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 553 AAGCCTGTGACGTACCAAGGCTACTATGCTGCTCACTCCAACTGGATATAACAGCCAC 612
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 820 GGTGTAGTAAAGGAGTTTCCTTCGTATCATTTTATTCACATCGTTTGGAAAAGACTTAC 646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 613 AACGAAGACTATACATCGTTGAGCAGTATGAAA 646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 880 GTAGAAGACGGGGGTTCTGTTGAACAGCATGAGA 913
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-09-626-581D-64
; Sequence 64, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
; LENGTH: 1079

; TYPE: DNA
; ORGANISM: Renilla muelleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; OTHER INFORMATION:
US-09-626-581D-64

Query Match      19.6%; Score 133.2; DB 4; Length 1079;
Best Local Similarity 50.8%; Pred. No. 4.2e-36;
Matches 321; Conservative 0; Mismatches 313; Indels 0; Gaps 0;

Qy 13 AAGAATGTTATCAAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGAACGGTCAATGGG 72
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 280 AACACTTGTGTTACAAAGTAATGTCGTATAAGTAATCTCGAAGGAATTTGTAACAC 339
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 73 CACGAGTTTGAATAGAACGCGAAGGAGGAGGCCATACGAAGGCCAACATACCGTA 132
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 340 CATGTTTTCACAAATGAGGTTTCGGCAAGGAATATTTTATTCGGCAATCACTGGTT 399
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 133 AAGCTTAAAGTAAACAAAGGGGGACCTTTGCCATTTGGTGGGATATTTTGTCAACCAA 192
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 400 CAGATTCTGTGCACGAAGGGGCCCACTGCTTTTGCATTTGATATTTGTGCACCACT 459
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 193 TTTTCAGTATGAACAGGATATATGTCAGCACCTGCCGACATACCAAGCTATAAAAG 252
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 460 TTTCAATATGCAACCGTACTTTTCACGAATATCCGAATGATATATCAGATTATTTATA 519
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 253 CTGTCTATTCCTGAAGGATTTAAATGGGAAGGGTTCATGAACCTTTGAAGACGGTGGCGTC 312
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 520 CAATCATTTCCAGCAGAGTTTATGATGAACGAACATTCAGTTACGAAGATGGCGACTT 579
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 313 GTTACTGTAAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTCAAGGTCAGGTC 372
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 580 GTTCAATTCGTTCCAGAGATGATGATGAACGAACATTCAGTTACGAAGATGGCGACTT 639
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 373 ATTGGCGTGAACCTTCCCTCGATGGCGTGTGAAGGAGAGATTCATAAGGCT 492
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 700 CCTTCATTGAAGCCATGATCATGATGAATGAGCTGTTGGTCGGGAAGTAATCTTGTG 759
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 493 CTGAAGCTGAAGACGGTGGTTCATTAAGTGAATTCAGGATTTTCAATGCGCAAG 552
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 760 TATAAATTAACCTCTGGGAATATTAATCATGTACATGAAACATTAATGAAGTCGAA 819
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 553 AAGCCTGTGACGTACCAAGGCTACTATGCTGCTCACTCCAACTGGATATAACAGCCAC 612
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 820 GGTGTAGTAAAGGAGTTTCCTTCGTATCATTTTATTCACATCGTTTGGAAAAGACTTAC 646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 613 AACGAAGACTATACATCGTTGAGCAGTATGAAA 646
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 880 GTAGAAGACGGGGGTTCTGTTGAACAGCATGAGA 913
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-415-765B-64
; Sequence 64, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 64
```

LENGTH: 1079
TYPE: DNA
ORGANISM: Renilla muelleri
FEATURE:
NAME/KEY: CDS
LOCATION: (259)..(975)
OTHER INFORMATION:
US-09-415-765B-64

Query Match 19.6%; Score 133.2; DB 4; Length 1079;
Best Local Similarity 50.6%; Pred. No. 4.2e-36;
Matches 321; Conservative 0; Mismatches 313; Indels 0; Gaps 0;

```
Qy 13 AAGATGTTATCAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGAACGTCATGGG 72
Db 280 AACACTTGTGTTTACAGAGTAATGTCGTATAAAGTAATCTGGAAGGAATTTGTAACAAC 339
Qy 73 CACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGCCCAATACCGTA 132
Db 340 CATGTTTTCATAGGAGGTTTCGCGCAAGGGAATATTTATTCGGCAATCACTGGTT 399
Qy 133 AAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTTGGGATATTTTGTCAACCAA 192
Db 400 CAGATTGCTGTCAGAAAGGGGGCCCACTGCTTTTGCATTTGATATTTGTGCACCACT 459
Qy 193 TTTCAGTATGAACCAAGGTATATGTCAGCACCTCGCGACATACCAAGACTATAAAAG 252
Db 460 TTTCAATATGCGAACCCGCTACTTTCACGAAATATCCGAATGATATATCAGATTATTTATA 519
Qy 253 CTGTCAATTTCTGAGGAGTTTAAATGGGAAGGGTCAATGAACTTTGAAGACGGTGGCGTC 312
Db 520 CAATCAATTTCCAGAGGATTTATGTAAGAACAACTTTCGTCTACAGAGTGGGACTT 579
Qy 313 GTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCTATCAAGGTCAGGTTTC 372
Db 580 GTTGAATTCGTTTTCAGATATAAATTAATAGAAAGACAAGTTTCGTCTACAGAGTGGAAATAC 639
Qy 433 GCCAGCACTGAGCGTTTGTATCTCTGATGGGCTGTTGAAAGGAGAGATTTCATAAGGCT 492
Db 700 CCTTCATTTGAAGCCATGTACATGAATATGCGCTTGGTGGCGAAGTAATTTCTTGT 759
Qy 493 CTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAATTTCAAAAGTATTTACATGGCAAG 552
Db 760 TATAAATCAACTCTGGAAATATTTATTCATGTACATGAAACATTAATGAAGTGGAA 819
Qy 553 AAGCCTGTGACGTACAGGGTACTATGTTGACTCCAACTGGGATATAACAGCCAC 612
Db 820 GGTGTAGTAAAGGAGTTTCTTCGTATCATTTTATTCACATCGTTTGGAAAGACTTAC 879
Qy 613 AACGAAGACTATACAATCGTTGAGCAGTATGAAA 646
Db 880 GTAGAAGACGGGGGTTTCGTTGAACAGCATGAGA 913
```

RESULT 12

US-09-626-580C-64
Sequence 64, Application US/09626580C
Patent No. 6562617
GENERAL INFORMATION:
APPLICANT: Anderson, David
APPLICANT: Bogenberger, Jakob M.
APPLICANT: Pele, Beau R.
TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
FILE REFERENCE: A-66900-2/RMS/AMS
CURRENT APPLICATION NUMBER: US/09/626,580C
CURRENT FILING DATE: 2000-07-27
PRIORITY FILING DATE: US 09/415,765
PRIORITY FILING DATE: 1999-10-08
PRIORITY FILING DATE: 1999-03-26

PRIOR FILING DATE: 1998-10-08
NUMBER OF SEQ ID NOS: 65
SOFTWARE: Patent in version 3.1
SEQ ID NO 64
LENGTH: 1079
TYPE: DNA
ORGANISM: Renilla muelleri
FEATURE:
NAME/KEY: CDS
LOCATION: (259)..(975)
OTHER INFORMATION:
US-09-626-580C-64

```
Query Match 19.6%; Score 133.2; DB 4; Length 1079;  
Best Local Similarity 50.6%; Pred. No. 4.2e-36;  
Matches 321; Conservative 0; Mismatches 313; Indels 0; Gaps 0;
```

Qy 13 AAGATGTTATCAGGAGTTTCATGAGTTTAAAGTTTCGCATGGAAGAACGTCATGGG 72
Db 280 AACACTTGTGTTTACAGAGTAATGTCGTATAAAGTAATCTGGAAGGAATTTGTAACAAC 339
Qy 73 CACGAGTTTGAATAGAGGCGAAGGAGAGGGAGGCCATACGAAGGCCCAATACCGTA 132
Db 340 CATGTTTTCATAGGAGGTTTCGCGCAAGGGAATATTTATTCGGCAATCACTGGTT 399
Qy 133 AAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTTGGGATATTTTGTCAACCAA 192
Db 400 CAGATTGCTGTCAGAAAGGGGGCCCACTGCTTTTGCATTTGATATTTGTGCACCACT 459
Qy 193 TTTCAGTATGAACCAAGGTATATGTCAGCACCTCGCGACATACCAAGACTATAAAAG 252
Db 460 TTTCAATATGCGAACCCGCTACTTTCACGAAATATCCGAATGATATATCAGATTATTTATA 519
Qy 253 CTGTCAATTTCTGAGGAGTTTAAATGGGAAGGGTCAATGAACTTTGAAGACGGTGGCGTC 312
Db 520 CAATCAATTTCCAGAGGATTTATGTAAGAACAACTTTCGTCTACAGAGTGGGACTT 579
Qy 313 GTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCTATCAAGGTCAGGTTTC 372
Db 580 GTTGAATTCGTTTTCAGATATAAATTAATAGAAAGACAAGTTTCGTCTACAGAGTGGAAATAC 639
Qy 373 ATTGGCGTGAACCTTCTTCGATGGACCTGTTATGCAAAAGACAAATGGGCTGGGAA 432
Db 640 AAGGTTAGTAATCTCCAGATGATGGTCCCGCTCATGCAAGAGACTATCTTAGGAATAGAG 699
Qy 433 GCCAGCACTGAGCGTTTGTATCTCTGATGGGCTGTTGAAAGGAGAGATTTCATAAGGCT 492
Db 700 CCTTCATTTGAAGCCATGTACATGAATATGCGCTTGGTGGCGAAGTAATTTCTTGT 759
Qy 493 CTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAATTTCAAAAGTATTTACATGGCAAG 552
Db 760 TATAAATCAACTCTGGAAATATTTATTCATGTACATGAAACATTAATGAAGTGGAA 819
Qy 553 AAGCCTGTGACGTACAGGGTACTATGTTGACTCCAACTGGGATATAACAGCCAC 612
Db 820 GGTGTAGTAAAGGAGTTTCTTCGTATCATTTTATTCACATCGTTTGGAAAGACTTAC 879
Qy 613 AACGAAGACTATACAATCGTTGAGCAGTATGAAA 646
Db 880 GTAGAAGACGGGGGTTTCGTTGAACAGCATGAGA 913

RESULT 13

US-09-277-716-15
Sequence 15, Application US/09277716A
Patent No. 6232107
GENERAL INFORMATION:
APPLICANT: Bryan, Bruce
APPLICANT: Szent-Gyorgyi, Christopher
APPLICANT: PROLUME, LTD.
TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
CURRENT APPLICATION NUMBER: US/09/277,716A
CURRENT FILING DATE: 1999-03-26

; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1079
; TYPE: DNA
; ORGANISM: Renilla mulleri
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (259)..(975)
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-277-716-15

Query Match 19.6%; Score 133.2; DB 3; Length 1085;
Best Local Similarity 50.6%; Pred. No. 4.2e-36;
Matches 321; Conservative 0; Mismatches 313; Indels 0; Gaps 0;
QY 13 AAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGAACGGTCAATGGG 72
Db 280 AACACTTTGTTTACAAAGAGTAATCGTGTATAAAGTAAATCTGGAAGGAATTTGTAACAAC 339
QY 73 CACGAGTTTGAATAGGAAGGAGAGAGGGAGGCCATACGAAGGCCACATACCGTA 132
Db 340 CATGTTTTTAAATGGAAGGGTTCGGCAAAAGGAATATTTATTCGGCAATCAACTGGTT 399
QY 133 AAGCTTAAGGTAACCAAGGGGGACCTTTGCACTTTGCTGGGATATTTTGTACCAACAA 192
Db 400 CAGATTGCTGTCAGAAAGGGGCCCTGCTTTTGCAATTTGATTTGTCTACAGCT 459
QY 193 TTTCAAGTATGAACCAAGGTTATATGTCAAGCACCTTCCCGACATACCAGACTATAAAG 252
Db 460 TTTCATATGTCACCAACCGTACTTTCACGAATATCCGAATGATATATCAGATTATTTTATA 519
QY 253 CTGTCATTTCTGAAGGATTTAAATGGGAAGGGTTCATGACTTTGAAGCGGTGGCGTC 312
Db 520 CAATCATTTTCCAGCAGGATTTATGATATGAAGCAACATTAAGTTACGAAGATGGCGACTT 579
QY 313 GTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTACAAGTCAAGTTC 372
Db 580 GTTCAAAATTCGTCAGATATAAATTTAATAGAACAGAGTTCGTCACAGTGGATAC 639
QY 373 ATTGGGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAACAAATGGGCTGGAA 432
Db 640 AAAGGTAGTAACCTTCCAGATGATGGTCCCGTCATGCAAGAACTATCTTAGGAATAGAG 699
QY 433 GCCAGCACTGAGCGTTGTATCCTCGTATGGCGGTGTTGAAGGAGAGATTCAAGGCT 492
Db 700 CTTTCATTTGAAGCCATGTACATGAATAAGGCGCTTTGGTCGGCAAGTAATCTTTGTC 759
QY 493 CTGAAGCTGAAGACGGTGGTTCATTACCTAGTTGAAATTCAAAGTATTTATCATGGCAAG 552
Db 760 TATAAATCTAACTCTGGGAATATTTATCTGTACATGAATAAATTAATGAAGTCCAAA 819
QY 553 AAGCCTGTGACGTACCAAGGTTACTATGTTGACTCCAACTCCGATATTAACAAAGCCAC 612
Db 820 GGTGTAGTAAAGGAGTTTCCCTGCTATCATTTTATTAACATCATCTGTTGGAAGAACCTTAC 879
QY 613 AACGAAGCTATACATCTGTTGACAGTATGAAAGAACCGAGGAC 646
Db 880 GTAGAAGACGGGGGTTGTTGTAACAGCATGAGA 913

RESULT 14

US-09-277-716-30
; Sequence 30, Application US/09277716A
; Patent No. 6232107
; GENERAL INFORMATION:

; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE
; CURRENT APPLICATION NUMBER: US/09/277,716A
; CURRENT FILING DATE: 1999-03-26
; EARLIER APPLICATION NUMBER: 60/102,939
; EARLIER FILING DATE: 1998-10-01
; EARLIER APPLICATION NUMBER: 60/089,367
; EARLIER FILING DATE: 1998-06-15
; EARLIER APPLICATION NUMBER: 60/079,624
; EARLIER FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; FEATURE:
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-277-716-30

Query Match 19.5%; Score 132.4; DB 3; Length 1104;
Best Local Similarity 50.3%; Pred. No. 8.1e-36;
Matches 325; Conservative 0; Mismatches 321; Indels 0; Gaps 0;
QY 13 AAGATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGAACGGTCAATGGG 72
Db 55 AACACTGGACTGAAGAGATTAATGTCGGCAAAAGCTAGCGTTGAAGGAATCGTGAACAAT 114
QY 73 CACGAGTTTGAATAGGAAGGAGAGAGGGAGGCCATACGAAGGCCACATACCGTA 132
Db 115 CACGTTTTTTCATGGAAGGATTTGAAAAGGCAATGTATTTTGGAAACCAATTTGATG 174
QY 133 AAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTCAACCAA 192
Db 175 CAAATCCGGGTTACAAAGGGAGTCCGTTGCCATTCGTTCCATATTTGTTCCATAGCT 234
QY 193 TTTCAAGTATGAAGCAAGGTTATATGTCAGCACCTTCCCGACATACACAGACTATAAAG 252
Db 235 TTCCAATACGGGAATCGCACTTTTACGAAATACCCAGACGACATTCGGGACTACTTTGTT 294
QY 253 CTGTCATTTCTGAAGGATTTAAATGGGAAGGGTTCATGAACTTTGAAGACGGTGGCGTC 312
Db 295 CAATCATTTCCCGCTGGATTTTCTACGAAGAAATCTACGCTTTGAAGATGGGCCATT 354
QY 313 GTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTACAAGTCAAGTTC 372
Db 355 GTTGACATTCGTTTCAGATATAAGTTTAGAAGATGATAAGTTCCCATCAAAAGTGGAGTAT 414
QY 373 ATTGGGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAACAAATGGGCTGGAA 432
Db 415 AGAGGCAACGGTTTCCCTAGTAACGGACCGCTGATGCAAAAGCCATCTCTCGCATGGAG 474
QY 433 GCCAGCACTGAGCGTTTGTATCCTCGTATGGCGGTGTTGAAGGAGAGATTTCATAAGCT 492
Db 475 CCATCGTTTGAAGTGGTCTACATGAACAGCGCGGCTTCGTGGGCGGAAGTAGATCTGTT 534
QY 493 CTGAAGCTGAAGACGGTGGTTCATTACCTAGTTGAAATTCAAAAGTATTTTACATGGCAAG 552
Db 535 TACAACCTCGAGTCAGGAACTATTACTCGTGGCCACATGAACAGCTTTTACAGATCCAA 594
QY 553 AAGCCTGTGACGTACCAAGGTTACTACTATGTTGATCTCCAACTGGATATAACAGCCAC 612
Db 595 GGTGGAGTGAAGAAATTCGCGGAATATCACTTTATCCATCATCTGTTGGGAAACCTAC 654
QY 613 AACGAAGCTATACAACTCGTTGACAGTATGAAAGAACCGAGGAC 658
Db 655 GTGAAGAGGAAGCTTCGTGGGAACAAACACGAGACGGCCATTGAC 700

RESULT 15

US-09-609-161B-30
; Sequence 30, Application US/09609161B
; Patent No. 6436682
; GENERAL INFORMATION:
; APPLICANT: Bryan, Bruce
; APPLICANT: Szent-Gyorgyi, Christopher
; APPLICANT: PROLUME, LTD.
; TITLE OF INVENTION: LUCIFERASES, FLUORESCENT PROTEINS, NUCLEIC ACIDS ENCODING THE LUC
; TITLE OF INVENTION: AND FLUORESCENT PROTEINS AND THE USE THEREOF IN DIAGNOSTICS, HIG
; TITLE OF INVENTION: SCREENING AND NOVELTY ITEMS
; FILE REFERENCE: 24729-121B
; CURRENT APPLICATION NUMBER: US/09/609,161B
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 09/277,716
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/102,939
; PRIOR FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1104
; TYPE: DNA
; ORGANISM: Ptilosarcus gurneyi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (34)..(747)
; OTHER INFORMATION: Ptilosarcus Green Fluorescent Protein (GFP) (insert A)
US-09-609-161B-30

Query Match 19.5%; Score 132.4; DB 3; Length 1104;
Best Local Similarity 50.3%; Pred. No. 8.1e-36;
Matches 325; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

Qy	13	AAGAATGTTATCAAGGAGTTTCATGAGTTTAAAGTTCGCATGGAAGGAAACGGTCAATGGG	72
Db	55	AACACTGGACTGAAGAGATTATGTCGGCAAAAGCTAGCGTTGAAGGATCGTGAACAAT	114
Qy	73	CACGAGTTTGAATAGAGGGAAGAGAGAGGAGGAGGCCATACGAAGGCCACATACCCGTA	132
Db	115	CACGTTTTTCCATGGAAGGATTTGGAAAGGCAATGATATTTTGGAAACCAATTGATG	174
Qy	133	AAGCTTAAGTTAACAGGGGGACCTTGCCATTTGCTTGGGATATTTTGTACACCAA	192
Db	175	CAATCCGGGTACAAAGGGAGGTCGGTTCGCAATTCGATATTTGTCATAGCT	234
Qy	193	TTTCAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCAGACATACCAGACTATAAAAG	252
Db	235	TTCCAAATACGGGATCGCACTTTCCAGCAATACCCAGACACATTCGGACTACTTTGTT	294
Qy	253	CTGTCAATTTCTGAAGGATTTAAATGGGAAGGTCATGAACCTTTGAAGACGGTGGCGTC	312
Db	295	CAATCAATCCCGGCTGGATTTTCTACGAAAGAAATCTACGCTTTGAAGATGGGCCATT	354
Qy	313	GTTACTGTACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTACAGGTCAAGTTC	372
Db	355	GTTGACATTCGTTCCAGATATAAGTTTGAAGATGATAAAGTTCCACTACAAAGTGGAGTAT	414
Qy	373	ATTGGCGTGAATTTCTTCGATGGACCTGTTATGCAAAAGAGACAAATGGGCTGGAA	432
Db	415	AGAGGCAACGGTTTCCCTAGTAGCGGACCCCGTGATGCAAAAGCCATCTCTGGCATGAG	474
Qy	433	GCCAGCACTGAGCGTTTGTATCTCGTATGGCGGTGTTGAAAGGAGAGATTTCAATAGGCT	492
Db	475	CCATCGTTTGAGGTGGTCTACATGAACAGCGGCGTCTCTGGTGGCGAAGTAGATCTCGTT	534
Qy	493	CTGAAGCTGAAGACGGTGGTTCATTACCTAGTTGCAATTTCAAAAGTATTTTACATGGCAAAG	552

Db	535	TACAAACTCGAGTCAGGAACTATTACTCGTGCCACATGAAACGTTTTTACAGATCCAA	594
Qy	553	AAGCCTGTGCAGCTACCCAGGGTACTACTATGTTGACTCCAAACTGGATATACAAAGCCAC	612
Db	595	GGTGGAGTGAAGAATTTCCCGGAATATCACTTTTATCCATCATCGTCTCTGGAGAAAACCTAC	654
Qy	613	AACGAAAGACTATACAATCGTTGAGCAGTATGAAGAACCAGGGAC	658
Db	655	GTGGAAGAAGGAAGCTTCGTGGAAACAACACGAGACGGCCATTGAC	700

Search completed: July 1, 2005, 18:23:04
Job time : 173 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 30, 2005, 15:50:30 ; Search time 37.5 Seconds
(without alignments)
3479.198 Million cell updates/sec

Title: US-10-006-922a-11

Perfect score: 1215

Sequence: 1 atgaggtcttccaaagtgt.....gccaccatctgttcctttaa 678

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Xgapext 0.5	
Xgapop 6.0	Xgapext 7.0	
Deiop 6.0	Delext 7.0	

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 566832

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgm2_1/USPTO_spool_p/US10006922/runat_30062005_110552_9435/app_query.fasta_1.839
-DB=PIR -QPMT=fastan -SUFF=1rpr -MINMATCH=0.1 -LOPECL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=spect -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USR=US10006922 @CGN 1 1 63 @runat_30062005_110552_9435 -NCPU=6 -ICPU=3
-NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DSLOP=6 -DELEXI=7

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213.5	17.6	238	1 JQ1514	green-fluorescent
2	93.5	7.7	26926	1 I38344	titin, cardiac mus
3	92	7.6	1873	2 T30944	surface protein pr
4	91	7.5	458	1 A42386	hsp 90-binding pr
5	90	7.4	568	2 T06489	probable peptidylp
6	89	7.3	1433	1 A36734	bacillopeptidase F
7	87.5	7.2	1116	2 B70476	hypothetical prote
8	87	7.3	187	2 T01852	probable blue copp
9	87	7.2	1484	2 C37196	probable membrane
10	85.5	7.0	456	2 G69397	signal-transducing
11	85.5	7.2	1206	2 E86445	hypothetical prote
12	85	7.0	271	2 F69442	hypothetical prote
13	84.5	7.0	679	2 A40351	adhesion-type prot
14	84.5	7.0	725	1 IJMSNG	neural cell adhesi

15	84.5	7.0	2004	2	AC0314	probable membrane
16	84	7.0	839	2	T04859	extensin homolog F
17	83.5	6.9	1616	2	T17884	S-layer protein -
18	83.5	6.9	268	2	E90276	conserved hypochet
19	83.5	6.9	340	2	E69544	hypothetical prote
20	83.5	6.9	374	2	T06245	gibberellin 3 beta
21	83.5	6.9	374	2	T06244	gibberellin 3 beta
22	83	6.8	373	2	T50605	hypothetical prote
23	83	6.8	602	2	E81439	threonine-tRNA lig
24	83	6.8	15281	2	S41309	cyclosporin synthe
25	82.5	6.8	559	1	S55383	peptidylprolyl iso
26	82.5	6.9	589	2	E71251	probable oligoendo
27	82.5	6.8	680	2	S17982	Kallmann syndrome
28	82.5	6.8	751	1	I49497	transforming prote
29	82.5	6.8	931	2	T32919	hypothetical prote
30	82	6.7	341	2	B53125	restriction enzyme
31	81.5	6.7	292	2	C69106	glucose-1-phosphat
32	81.5	6.7	296	2	S17989	La 4.1 protein - h
33	81.5	6.7	551	1	S72485	peptidylprolyl iso
34	81.5	6.7	629	2	C64180	hypothetical prote
35	81.5	6.7	6805	2	S20901	titin - rabbit (fr
36	81	6.7	346	2	S77025	nitrilase (EC 3.5.
37	81	6.7	399	2	T40831	probable guanine n
38	81	6.7	587	1	E69171	phosphoesterase-re
39	80.5	6.6	404	1	S03849	ribonucleoprotein
40	80.5	6.6	415	1	JC1494	ig mu chain C regi
41	80.5	6.7	450	1	MHDG	hypothetical prote
42	80.5	6.6	862	2	F75116	phosphoribosylamin
43	80.5	6.6	1003	1	AJCHPR	neural cell adhesi
44	80.5	6.6	1115	1	IJMSNL	choline binding pr
45	80	6.6	621	2	A95250	

ALIGNMENTS

RESULT 1

JQ1514
green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)
C/Species: Aequorea victoria
C/Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 09-Jul-2004
C/Accession: JQ1514; JQ1514; PQ0335; S48693; S51330; S51331
R/Prasher, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.
Gene 111, 229-233, 1992
A/Title: Primary structure of the Aequorea victoria green-fluorescent protein.
A/Reference number: JQ1514; MUID:92175527; PMID:1347277
A/Accession: JQ1514
A/Molecule type: DNA
A/Residues: 1-107,'S', 109-238 <PRA1>
A/Cross-references: UNIPROT:P4212; UNIPROT:Q17106; UNIPROT:Q17105; GB:M62654; NID:g1556
A/Accession: JQ1514
A/Molecule type: mRNA
A/Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>
A/Cross-references: GB:M62653; NID:g155660; PIDN:AAA27721.1; PID:g155661
A/Accession: PQ0335
A/Molecule type: protein
A/Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>
R/Inouye, S.; Tsuji, F.I.
FEBS Lett. 351, 211-214, 1994
A/Title: Evidence for redox forms of the Aequorea green fluorescent protein.
A/Reference number: S48693; MUID:94364470; PMID:8082767
A/Accession: S48693
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>
A/Cross-references: GB:I29345; NID:g606383; PIDN:AAA58246.1; PID:g606384
R/Watkins, J.N.; Campbell, A.K.
submitted to the EMBL Data Library, January 1995
A/Reference number: S51330
A/Accession: S51330
A/Molecule type: mRNA
A/Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'R',
A/Cross-references: EMBL:X83959; NID:g634008; PIDN:CAA58789.1; PID:g634009
A/Experimental source: clone gfp1

submitted to the Brookhaven Protein Data Bank, February 1996

A;Reference number: A66736; PDB:1IIT
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341

R;Fuhi, M.; Pastore, A.

submitted to the Brookhaven Protein Data Bank, August 1996

A;Reference number: A66201; PDB:1NCT

A;Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155

C;Genetics:

A;Gene: GDB:TTN

A;Cross-references: GDB:127867; OMIM:188840

A;Map position: 2q31-2q32

C;Function:

A;Description: structural protein forming filaments in striated muscle
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco
structural protein

F;24752-25008/Domain: protein kinase homology <KIN>

F;84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted

F;16780,16976,17579,17602,17667,17681,17845,17899,18121,18189,18209,18336,18670,18680,18
21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Alignment Scores:

Pred. No.:	2.81	Length:	26926
Score:	93.50	Matches:	51
Percent Similarity:	34.55%	Conservative:	34
Best Local Similarity:	20.73%	Mismatches:	84
Query Match:	7.70%	Indels:	77
DB:	1	Gaps:	13

US-10-006-922A-11 (1-678) x I38344 (1-26926)

QY	43	AAGTTCGCATGGAAGAACCGTCAATGGGCACGATTTGAA-----ATA 87
DB	23493	ArgTyrArgSerThrGlyLeuThrGluGlyLeuGluThrGluHisArgValThrAlaIle 23512
QY	88	GAAGCGAAGGAGGAGGAGCCATACGAGCCACCAATACCGTA-----132
DB	23513	AsnAlaArgGlySerGlyLysProSerArgProSerLysProIleValAlaMetAspPro 23532
QY	133	-----AAGCTTAAGTAACCAAGGGGACCTTG 162
DB	23533	IleAlaProGlyLysProGlnAsnProArgValThrAspThrArgThrSerVal 23552
QY	163	CAATTCCTGGGATATTTGTGCACCAATTTCCAGTATGGAAGCAAGTA-----TAT 216
DB	23553	SerLeuAlaTrpSerVal-----ProGluAspGluGlySerLysValThrGlyTyr 23570
QY	217	GTC-----AAGCAC-----CCTGCC 231
DB	23571	LeuIleGluMetGlnLysValAspGlnHisGluTrpThrLysCysAsnThrProThr 23590
QY	232	GACATACCACTATAAAGCTGCTCATTTCTGCAAGGATTTAAATGGGA-----AGGTC 288
DB	23591	LysIleArgGluTyrThrLeuThrHisLeuProGlnGlyAlaGluTyrArgPheArgVal 23610
QY	289	ATGAACCTTTGAAGACGGT-----GGCGTCGTTACTGTA 321
DB	23611	LeuAlaCysAsnAlaGlyGlyProGlyGluProAlaGluValProGlyThrValLysVal 23630
QY	322	ACCCAG-----GATTCACAGTTTGCAGGATGGCTGTTTC 354
DB	23631	ThrGluMetLeuGluTyrProAspTyrGluLeuAspGluArgTyrGlnGluGlyIlePhe 23650
QY	355	ATCTACAGGTCAGTTTCATTTGGCGTGAACCTTCTCCGATGA-----CCTGTT 405
DB	23651	ValArgGlnGlyGlyValIleArgLeuThrIleProIleLysGlyLysProPheProIle 23670
QY	406	ATGCAAAAGAACAAATGGGCTGGGAAGCCAGCACTGAGCGTTGTATCTCGTGATGGC 465
DB	23671	CysLys-----TrpThrLysGluGlyGlnAspIleSerLysArgAlaMet 23685

QY	466	GTGTTGAAGGAGAGAGATTCTAAGGCTCTCAAGCTGAAA-----GACGGT 510
DB	23686	IleAlaThrSerGluThrHisThrGluLeuValIleLysGluAlaAspArgGlyAspSer 23705
QY	511	GGTCATTACTAGTCTGAATTC-----AAAAGTATTTACATGGCAAG 552
DB	23706	GlyThrTyrAspLeuValLeuGluAsnLysCysGlyLysAlaValTyrIleLysVal 23725
QY	553	AAGCCTGTGCAGCTACCA 570
DB	23726	ArgValIleGlySerPro 23731

RESULT 3
T30944

surface protein precursor - Enterococcus faecalis

C;Species: Enterococcus faecalis

C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

C;Accession: T30944

R;Shankar, V.; Baghdadyan, A.S.; Huycke, M.M.; Lindahl, G.; Gilmore, M.S.

Infect. Immun. 67, 193-200, 1999

A;Title: Infection-derived Enterococcus faecalis strains are enriched in esp, a gene enc

A;Reference number: Z20943; MUID:99081742; PMID:9864215

A;Accession: T30944

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-1873 <SHA>

A;Cross-references: UNIPROT:Q9Z4N7; EMBL:AF034779; NID:g3873186; PID:g3873187; PIDN:AADO

Alignment Scores:			
Pred. No.:	2.17	Length:	1873
Score:	92.00	Matches:	46
Percent Similarity:	35.75%	Conservative:	33
Best Local Similarity:	20.81%	Mismatches:	56
Query Match:	7.57%	Indels:	86
DB:	2	Gaps:	13

US-10-006-922A-11 (1-678) x T30944 (1-1873)

QY	88	GAAGCGAAGGAGGAGGAGCCATACGAGCCACCAATACCGTAAAGCTTAAGTAAACC 147
DB	708	LysGlyValGlyGluSerGluProIleThrGly-----ThrIleProIleLysVal 725
QY	148	AAGGGGGACCTTGGCCATTTGCTGGGATATTTGTCCACCAATTTTCAGTAGTGAAGC 207
DB	726	-----TyrGlnAspGlySer 730
QY	208	-----AAGGTATATGTCAAGCACCTGCC 231
DB	731	ValGlyThrThrAspLeuAlaValThrValSerLysAsnIleTyrGluAsnProGlyGlu 750
QY	232	GACATACCA---GACTATATAAAGCTGCTCATTTCTCCTGAAGGATTTAAATGGAAAGGTC 288
DB	751	AsnIleProAlaGlyTyrHisLysValThrPheThrAlaGly-----764
QY	289	ATGAACCTTTGAAGACGGTGGCGTCTGTACTGTACCCAGGATTCACGTTTCAGAGATGGC 348
DB	765	-----GluGlyThrSerIleGluSerGly 772
QY	349	TGTTTCATCTACAGTCAAGTTCATTTGCGGTGNACTTTCTTCCGATGGA---CCTGTT 405
DB	773	ThrThrValPheAlaValLys---AspGlyValSerLeuProGluAspLysLeuProVal 791
QY	406	ATGCAAAAGAACAAATGGGCTGGGAGCCAGCACCTGAGCGTTGTATCTCGTGATGGC 465
DB	792	LeuLysAlaLys-----AspGlyTyrThrAspAlaLysTrpProGluGluAla 807
QY	466	GTG-----TTGAAAGAGAGAG-----ATTCAAGGCTCTGAAGCTGAAGAC 507
DB	808	ThrGlnProIleLysAlaAspThrGluPheValSerSerAlaThrLysLeuAspAep 827
QY	508	-----GGTGGTCATTACCTAGTGTGAATTC-----531

Db 165 ---AsnProLysAspLeuAspGluValThrValLysTyrGluAlaAArgLeuGluAspGly 183
Qy 349 TGTTCATCTACAGCTCAAGTTCATTGGCGTGAACCTTTCCTCCGAGGACCTGTTATG 408
Db 184 SerValValSerLysSerGluSerIle----- 192
Qy 409 CAAAGAAGACAATGGCGTGGAGCCAGCACTGAGCGTTTGATCTCTCGTGATGGCGTG 468
Db 193 -----GluPheSerVal----- 200
Qy 469 TTGAAGAGAGATTATAGCGCTCTGAAG---CTGAAGACGCGTGGTTCATTACTAGTT 525
Db 201 PheCysProAlaLeuSerLysAlaValLysThrMetLysLysGlyGluLysValLeuLeu 220
Qy 526 GAATTCAAAAGTATTAC-----ATGCAAGAAAGCGCTGTCAGCTACCGAGG 573
Db 221 ThrValLysProGlnTyrGlyPheGlyGluGlnGlyArgAlaAlaThrGluValGluGly 240
Qy 574 TACTACTATGTGACTCCAACTCGATATA----- 603
Db 241 AlaValProProAsnSerThrLeuHisLeuAspLeuGlnLeuValSerTrpLysThrLeu 260
Qy 604 ACAAGCCACACGAGACTATACATCGTT-----GAGCAGTAT 642
Db 261 ThrLeuLeuGlyAspAspLysArgIleLeuLysLysValLeuLysGluGlyGlyTyr 280
Qy 643 GAAAGAACCGAGGAGCCACCATC 667
Db 281 GluArg-ProAsnAspGlyAlaVal 288
RESULT 6
A36734
bacillopeptidase F (EC 3.4.21.-) precursor bpr [validated] - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text change 09-Jul-2004
C:Accession: A36734; A35131; A35750; B35750; S08223; JN0335; I39849; B69596; JU0084
R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 5520-5521, 1990
A:Reference number: A36734; MUID:90368623; PMID:2118514
A:Contents: erratum
A:Accession: A36734
A:Molecule type: DNA
A:Residues: 1-1433 <SLO>
A:Cross-references: UNIPROT:P16397; GB:M29035; NID:g143307; PIDN:AA62679.1; PID:g143308
R:Sloma, A.; Rufo Jr., G.A.; Rudolph, C.F.; Sullivan, B.J.; Theriault, K.A.; Pero, J.
J. Bacteriol. 172, 1470-1477, 1990
A:Title: Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning
A:Reference number: A35131; MUID:90170864; PMID:2106512
A:Accession: A35131
A:Molecule type: DNA
A:Residues: 1-365, 'S', 367-682, 'EIMP', 893, 'Q', 895-896 <SL2>
A:Cross-references: GB:M29035
A:Note: the authors translated the codon GAA for residue 545 as Leu
R:Wu, X.C.; Nathoo, S.; Pang, A.S.H.; Carne, T.; Wong, S.L.
J. Biol. Chem. 265, 6845-6850, 1990
A:Title: Cloning, genetic organization, and characterization of a structural gene encoding
A:Reference number: A35750; MUID:90216713; PMID:2108961
A:Accession: A35750
A:Molecule type: DNA
A:Residues: 1-392, 'V', 394-828, 'NIRTRLVSLKPCRSHKSV' <WUA>
A:Cross-references: GB:J05400; NID:g142607; PIDN:AAA83362.1; PID:g142609
A:Note: this sequence has been corrected
A:Accession: B35750
A:Molecule type: DNA
A:Residues: 876-935, 'CG' <WU2>
A:Cross-references: GB:J05400; NID:g142607; PIDN:AAA83363.1; PID:g1119197
A:Note: this sequence has been corrected
R:Masuda, E.S.; Anaguchi, H.; Sato, T.; Takeuchi, M.; Kobayashi, Y.
Nucleic Acids Res. 18, 657, 1990
A:Title: Nucleotide sequence of the sporulation gene spoIIIGA from Bacillus subtilis.
A:Reference number: S08223; MUID:90174995; PMID:2106671
A:Accession: S08223
A:Molecule type: DNA

A:Residues: 1410-1433 <MAS>
A:Cross-references: EMBL:X17344; NID:g40165; PIDN:CAA35224.1; PID:g809661
R:Kato, T.; Yamagata, Y.; Arai, T.; Ichishima, E.
Biosci. Biotechnol. Biochem. 56, 1166-1168, 1992
A:Title: Purification of a new extracellular 90-kDa serine proteinase with isoelectric pI
A:Reference number: JN0335; MUID:93005071; PMID:1368833
A:Accession: JN0335
A:Molecule type: protein
A:Residues: 195-218, 'A' <KAT>
A:Note: source of this material was Bacillus subtilis (natto)
R:Beall, B.; Lowe, M.; Lutkenhaus, J.
J. Bacteriol. 170, 4855-4864, 1988
A:Title: Cloning and characterization of Bacillus subtilis homologs of Escherichia coli
A:Reference number: I39849; MUID:89008108; PMID:3139638
A:Accession: I39849
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-211 <RES>
A:Cross-references: GB:M22630; NID:g142938; PIDN:AAA22458.1; PID:g551705
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, F.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yamamoto, K.; Yata, K.; Yoshida, K.
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: B69596
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1433 <KUN>
A:Cross-references: GB:299111; GB:Z99112; GB:AL009126; NID:g2633902; PIDN:CAB13404.1; PTI
A:Experimental source: strain 168
C:Genetics:
A:Gene: bpr, bpf
A:Map position: 135 (degrees)
C:Superfamily: bacillopeptidase F; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-30/Domain: signal sequence #status predicted <SIG>
F:31-194/Domain: propeptide #status predicted <PRO>
F:195-1433/Product: bacillopeptidase F #status experimental <MAT>
F:218-466/Domain: subtilisin homology <SET>
F:227,274,452/Active site: Asp, His, Ser #status predicted
Alignment Scores:
Pred. No.: 3.97 Length: 1433
Score: 89.00 Matches: 56
Percent Similarity: 35.19% Conservative: 26
Best Local Similarity: 24.03% Mismatches: 71
Query Match: 7.33% Indels: 80
DB: 1 Gaps: 14
US-10-006-922A-11 (1-678) x A36734 (1-1433)
Qy 49 CGCATGGAAGAACGGTCAATGGCAGCGTTCGATGAGGCGAGGAGGAGGAGGAGG 108
Db 521 LysAlaGluGlyGlnValSer-----ValGluGlyAspAspGlnGluPro 535
Qy 109 CCATACGAAGCCCAATACCGTAACGTTAAGTAAACCAAGGGGACCTTTGCCATT 168
Db 536 ProValTyrGlnHisGluLysVal---ThrGluAlaTyrGluGlySerLeuProLeu 554
Qy 169 GCT----- 171
Db 555 ThrLeuThrAlaGluAspAsnValSerValThrSerValLysLeuSerTyrLysLeuAsp 574

```

Qy 172 -----TGGGATATTTCACCAATTTTCAGTATGGAAGCAAGTATATGTCAAG 222
Db 575 GlnGlyGluTrpThrGluThrAlaLysArgIleSerGlyAspHisLeuLysGlyThr 594
Qy 223 CACCTCGCGACATACAGCATATAA-----AAGCTGTCAATTCCTGAAGGATTA 276
Db 595 TyrGlnAlaGluLeuProAspIleLysGlyThrLysLeuSer-----TyrLys 610
Qy 277 TGGGAAAGGTCATGAATTTTGAAGCGTGGC-----GTCGTTACT 318
Db 611 Trp-----MetIleHisAspPheGlyGlyHisValValSerSerAspValTyrAsp 627
Qy 319 GTAAACCCAGGATTCAGCTTTCCAGGATGGCTGTTTCATCTACAAGGTCAGGTTCC 372
Db 628 ValThrValLysProSerIleThrAlaGly-----TyrLysGlnAspPheGluThr 644
Qy 373 -----ATTGGCGTGAAC-----TTTCCTTCC 393
Db 645 AlaProGlyGlyTrpValAlaSerGlyThrAsnAsnAsnTrpGluTrpGlyValProSer 664
Qy 394 GATGACCTGTTATGCAAAAGAACATATGGCTGGGAGCCAGCACTGACGCTTTCTAT 453
Db 665 ThrGlyPro-----AsnThrAlaAlaSerGlyGluLysValTyr 677
Qy 454 CCTCGTGATGCGTGTGAAGAGGAG-----ATTTCAT 486
Db 678 GlyThrAsn-----LeuThrGlyAsnTyrAlaAsnSerAlaAsnMetAsnLeuValMet 695
Qy 487 AAGCTCTGAAGTGAAGAGCGGTGGTCATTACCTAGTGAATTCAAAGTATTTACATG 546
Db 696 ProProIleLysAlaProAspSerGlySerLeuPheLeuGlnPheLysSerTrpHisAsn 715
Qy 547 GCAAGAAGCTGTGCAGCTACCAAGGCTACTACTATGTT 585
Db 716 LeuGluAspAspPheAspTyr---GlyTyrValPheVal 727

RESULT 7
B70476
hypothetical protein aq_2054 - Aquifex aeolicus
C;Species: Aquifex aeolicus
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 09-Jul-2004
R;Accession: B70476
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
V.
Nature 392, 353-358, 1998
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A;Reference number: A70300; MUID:98196666; PMID:9537320
A;Accession: B70476
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1116 <AQF>
A;Cross-references: UNIPROT:O67838; GB:AE000770; NID:g2984274; PIDN:AAC07805.1; PID:g298
A;Experimental source: strain VF5
C;Genetics:
A;Gene: aq_2054

Alignment Scores:
Pred. No.: 5.23 Length: 1116
Score: 87.50 Matches: 56
Percent Similarity: 35.69% Conservative: 35
Best Local Similarity: 21.96% Mismatches: 103
Query Match: 7.20% Indels: 61
DB: 2 Gaps: 9

US-10-006-922a-11 (1-678) x B70476 (1-1116)

Qy 13 AAGAATGTTATCAAGGATTCATGAGGTTTAAAGTTCGATGGAAGCAAGTCAATGGG 72
Db 164 LysAspLeuLeuAspSerGluTyrPheSerAlaLysThrLysGlyGluIleLysArg 183
Qy 73 CACGAGTTTGAATAGAGCGCAAGGAGAG----- 102
Db 73 CACGAGTTTGAATAGAGCGCAAGGAGAG----- 102

```

```

Db 184 AsnThrGlyGluIleLeuAlaGluValGluIleLysGluIleArgLysGluAsnPheThr 203
Qy 103 -----GGGAGGCATACGAAGCCACATACCGTAAAGCTTAAGGTAAACCAAGGGGGA 156
Db 204 LeuSerGlyThrLysIleAsnAlaLysGlyThrIleAsnLeuProValLeuAspIleAsn 223
Qy 157 CCTTTGCCAFTTTCCTGGGATATTTTGTCCACCAATTTTCAGTATGGAAGC----- 207
Db 224 AlalysAlaPheValLysAspLeuIleValArgAsnLysAsnTyrGlySerIleGluGly 243
Qy 208 -----AAGGTATATGTCACGACCCCTGCCGCACATA 237
Db 244 IleValLysGlyAsnTyrGluLeuPheAspLysLeuPheLysGlyGluAlaValAsn 263
Qy 238 CCAGACTATAAAGCTGTCATTT-----CCTGAAGGA-----TTT 273
Db 264 ProGluGlyThrLysIleLysPheThrTyrAspValIleProGluGlyLeuLeuThrPhe 283
Qy 274 AAATGGGAAGG---GTCATGAACCTTTGAAGACGGTGGCGTCTGTTACTGTAAACCCAGGAT 330
Db 284 SerPheGluAsnLeuValValAspLysAsnThrLeuGlyIleAsnArgGluIleArgGly 303
Qy 331 TCAGTTTGCAGATGCGTGTTCATCTACAGGTCAAGTTCATTCATTCGGCGTGAATTCCT 390
Db 304 GluPheHisGlyAsnGlyLysValAspPheLysLysMetPheValLysValAsn----- 321
Qy 391 TCCGATGGACCTGTTATGCAAAAGAACATATGGGCTGGGAGCCAGCACTGAGCGTTTG 450
Db 322 -----AlaPheThrGluAsnLeu 327
Qy 451 TATCCTCGTATGCGCTGTTGAAGAGGAGATTCATAAGGCTCTGAAGCTGAAAGACGGT 510
Db 328 GluValIleAspLysLysPheLysGlyAspValLeuPheSerTyrAsnPheSerGlyAsn 347
Qy 511 GGTCAATACCTAGTTCGAATTCAAAAGTATTTACATGCGCAAGAGCCTGTGCGAGTACCA 570
Db 348 GlySerLeuAsnPheGluPheLysAsnSerGlyTyrAlaLys----- 361
Qy 571 GGGTACTACTATGTTGAC---TCCAACTGGATATATCAACGACCAACGAAAGACTATACA 627
Db 362 GlyAsnLeuIleAsnLysAsnLysLeuGluGlyGluPheSerPheAsnAspPhePro 381
Qy 628 ATCGTTGAGCAG-----TATGAAGAACCGAGGAGCGC 660
Db 382 ValValPheGlnAspPheAsnAlaTyrLeuSerGlyGluGlyLys 396

RESULT 8
T01852
probable blue copper-binding protein F9D12.16 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 26-Feb-1999 #sequence_revision 26-Feb-1999 #text_change 09-Jul-2004
C;Accession: T01852
R;Murray, J.; Langston, Y.; Ahrens, C.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of Arabidopsis thaliana F9D12.
A;Reference number: Z14444
A;Accession: T01852
A;Status: translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-187 <UNIPROT>
A;Cross-references: UNIPROT:O81500; EMBL:AF077407; NID:g3139339; PIDN:AAC26242.1; PID:g31
A;Experimental source: cultivar Columbia
C;Genetics:
A;Map position: 4
A;Introns: 56/1
A;Note: F9D12.16
C;Superfamily: stellacyanin
C;Keywords: copper; electron transfer; glycoprotein

Alignment Scores:
Pred. No.: 3.93 Length: 187
Score: 87.00 Matches: 35
Percent Similarity: 42.62% Conservative: 17

```


F;161-165/Region: heparin binding #status predicted
F;228-290/Domain: immunoglobulin homology <IMM3>
F;263-272/Region: NCAM binding #status predicted
F;323-388/Domain: immunoglobulin homology <IMM4>
F;420-482/Domain: immunoglobulin homology <IMM5>
F;519-596/Domain: fibronectin type III repeat homology <FN3A>
F;625-685/Domain: fibronectin type III repeat homology <FN3B>
F;41-96,139-189,235-288,330-386,427-480/Diulfide bonds: #status predicted
F;222,316,348,424,450,479/Binding site: carboxylate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 9.22 Length: 725
Score: 84.50 Matches: 37
Percent Similarity: 43.79% Conservative: 30
Best Local Similarity: 24.18% Mismatches: 57
Query Match: 6.95% Indels: 29
DB: 1 Gaps: 9

US-10-006-922A-11 (1-678) x IJMSNG (1-725)

```
QY 271 TTTAAATGGGAAGGTCATGAACCTTTGAAGACGGTGGCTGCTACTGTA----- 321
Db 548 PheThrTrpTyrAspAlaLysGluAlaAsnMetGluGlyIleValThrIleMetGlyLeu 567
QY 322 ACCCAGGATCCAGCTTTCGAGATGGCTGTTTCATCTACAAGGTCAAGTTCATTGGCGTG 381
Db 568 LysProGluThrThrTyrSerAsp-----ArgLeuAlaLeu 580
QY 382 AACTTTCTCCGATGACCTGTTATGCAAGAAGACAATGGCTGGGAAGCCAGCACT 441
Db 581 AsnGlyLysGlyLeuGlyGluIleMetGlnProSerGluSerLysThrGlnProValPro 600
QY 442 GAGCGTTTGTATCTCTGCTGATGGCTGTTGAAGAGAGATTCTAAG----- 489
Db 601 GluLeuSerAlaProLys-----LeuGluGlnMetGlyGluAspGlyAsnSer 617
QY 490 -----GCTCTGAAGCTGAAGACGGTGGT-----CATTACTAGTT 525
Db 618 IleLysValAsnLeuIleLysGlnAspAspGlyGlySerProIleArgHisTyrLeuVal 637
QY 526 GAATTCAAAAGTATTTACATGGCAAGAAAGCCT---GTGCAGCTACCA---GGGTACTAC 579
Db 638 LysTyrArgAlaLeuAlaSerGluTrpLysProGluIleArgLeuProSerGlySerHis 657
QY 580 TATCTTCACTCCAA---CTGGATATAACAGCCACCAACGAGACTATACAATCGTTGAG 636
Db 658 HisValMetLeuLysSerLeuAspTrpAsnAlaGluTyrGluValTyrValAlaGlu 677
QY 637 ---CAGTATGAAGAACCGAGGACCGCCACCATCTGTTCT 672
Db 678 AsnGlnGlnLysSerLysAlaAlaHisPheValPhe 690
```

RESULT 15

AC0314
probable membrane protein YP02573 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC0314
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AC0314
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2004 <KUR>
A:Cross-references: UNIPROT:Q8ZDJ2; GB:AL590842; PIDN:CAC91375.1; PID:g15980562; GSPDB:C
C:Genetics:
A:Gene: YP02573

Alignment Scores:

```
Pred. No.: 11.5 Length: 2004
Score: 84.50 Matches: 57
Percent Similarity: 31.52% Conservative: 30
Best Local Similarity: 20.65% Mismatches: 81
Query Match: 6.95% Indels: 108
DB: 2 Gaps: 9

US-10-006-922A-11 (1-678) x AC0314 (1-2004)

QY 55 GAAGGAACGGTCAATGGGACGAGTTTGAATAGAAGGGAAGGAGGAGGGAGCCATAC 114
Db 905 GluAsnArgHisAsnSerAspGlyPheGluThrGluGluGlnGluGlnThrThrAspLeu 924
QY 115 GAAGGACACAAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTGTCTGG 174
Db 925 GlnGlyMetAlaThrIleProLeuAspLeuLysSerTyrAlaAspAlaThrTyrGlnLeu 944
QY 175 GATATTTTGTACCAACAATTTCAG----- 198
Db 945 GlnLeuLeuSerGluAlaPheGluAlaGlyGlyArgSerValAlaAlaThrAlaArg 964
QY 199 -----TATGAACCAAGGTATATGTCAGCACCTGCCGACATA----- 237
Db 965 ValLeuValSerProTyrAspSerLeuValGlyValLysAlaAspGlyAspLeuSerTyr 984
QY 238 ----- 252
Db 985 IleAsnArgAspAlaValArgLysLeuAsnIleIleAlaValAspProSerLeuAsnLys 1004
QY 253 CTGTCATTTCTCTGAA-----GGATTTAAATGGGAAGGGTCATGAAC 294
Db 1005 IleAlaLeuProAspLeuSerLeuLeuGlnLysTyrIleSerValLeuThr 1024
QY 295 TTTGAAGACGGTGGCGTC----- 312
Db 1025 LysGlnAspSerGlyValTyrLysTyrGlnSerArgLeuLysGluGlnLeuValSerGlu 1044
QY 313 -----GTTACTGTAACCCAGGATTCCAGT 336
Db 1045 GlnProLeuGlnIleSerProThrGlyThrAspPheThrLeuValThrGln----- 1061
QY 337 TTCAGGATGGCTGTTTTCATCTACAAGGTCAAGTTCATTGGCGTGAATTTCTCTCCGAT 396
Db 1062 ---GlnProGlyAspPheIleLeuValLys-----AspSerGln 1074
QY 397 GGACCTGTTATGCAAAAG-----AAGACNAATGGGCTGGGAAGCCAGCAGCTGAGCTTTG 450
Db 1075 GlyGlnValLeuAsnArgIleSerTyrThrValAlaGlyAsnAlaAsnLeuThrArgSer 1094
QY 451 TATCCTCGTATGGCGTGTGAAAGGAGAGATTCTAAGGCTCTGAAGCTGAA----- 504
Db 1095 LeuAspArgAsnThrGluLeuLysLeuLysLeuAsnGlnAlaGluTyrLeuGlnGlyGlu 1114
QY 505 -----GACGCTGGTCAATTACCTAGTTGAA 528
Db 1115 GluIleGluIleAlaIleAsnAlaProTyrAlaGlySerGly-LeuIleThrIleGluLys 1134
QY 529 TTCAAAGATTTTACATGGCAAGAGAGCCCTGTGCAGCTACCAGGGGTACTACTATTGTGAC 588
Db 1134 sAspLysValTyrSerTrpGln----- 1141
QY 589 TCCAAACTGATATAACAGCCCAACAGAGACTATACAATCGTTG 634
Db 1142 -----TrpPheHisSerAspThrThrSerSerValGlnArgIle 1154
```

Search completed: June 30, 2005, 16:05:56
Job time : 66.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 30, 2005, 15:49:45 ; Search time 161.5 Seconds
(without alignments)
4299.563 Million cell updates/sec

Title: US-10-006-922a-11

Perfect score: 1215
Sequence: 1 atgaggcttccaagaatgt.....gccaccatctgtctctttaa 678

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 3224756

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-Q=/cgn2 1/USPTO spool p/US10006922/runat 30062005 110551 9411/app query.fasta 1.839
-DB=UniProt -QMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPCT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=spct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10006922 @CGN 1 1 244 @runat 30062005 110551 9411 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1214	99.9	225	Q9U6Y8	Q9U6Y8 discosoma s
2	1196	98.4	236	Q66ND7	Q66nd7 discosoma s
3	1193	98.2	225	Q6KFP8	Q6kf85 discosoma s
4	1191	98.0	236	Q66ND8	Q66nd8 discosoma s
5	1085.5	89.3	230	Q9GTJ7	Q9gtj7 discosoma s
6	804	66.2	221	Q66PV1	Q66pv1 acropora te
7	798	65.7	221	Q66PV8	Q66pv8 acropora ac
8	797	65.6	221	Q66PV0	Q66pv0 acropora mi
9	793	65.3	221	Q95P04	Q95p04 goniopora t
10	793	65.3	221	Q66PV9	Q66pv9 acropora hy
11	751	61.8	219	P83690	P83690 montipora e
12	745	61.3	227	Q66ND6	Q66nd6 discosoma s
13	729.5	60.0	232	Q9U6Y7	Q9uey7 discosoma s
14	690	56.8	225	Q963F5	Q963f5 montastraea
15	684	56.3	225	Q66ND3	Q66nd3 montastraea
16	681	56.0	225	Q95UA7	Q95ua7 montastraea

17	681	56.0	225	2	Q7Z0W4
18	679.5	55.9	227	2	Q7Z0W6
19	678.5	55.8	227	2	Q962P9
20	678.5	55.8	227	2	Q7Z0W8
21	678	55.8	225	2	Q7Z0W5
22	672.5	55.3	227	2	Q66ND2
23	670.5	55.2	227	2	Q66ND5
24	669.5	55.1	234	2	Q7Z0W7
25	668.5	55.0	234	2	Q8T5F2
26	667	54.9	266	2	Q9U6Y3
27	663	54.6	226	2	Q8T6U0
28	660.5	54.4	227	2	Q66ND4
29	655.5	54.0	226	2	Q66PUS
30	655.5	54.0	227	2	Q95V70
31	654.5	53.9	234	2	Q8MU47
32	652.5	53.7	225	2	Q7Z0W9
33	642.5	52.9	225	2	Q6USK3
34	639	52.6	224	2	Q8MU48
35	631.5	52.0	225	2	Q8T5F1
36	620.5	51.1	230	2	Q66FW1
37	610	50.2	223	2	Q6R8F5
38	608.5	50.1	231	2	Q8ISF8
39	599.5	49.3	225	2	Q8I6J8
40	580.5	47.8	219	2	Q6R8F4
41	579.5	47.7	259	2	Q8MMA2
42	574	47.2	228	2	Q9GFI6
43	573.5	47.2	235	2	Q8T5F0
44	570.5	47.0	231	2	Q8T6T8
45	566.5	46.6	239	2	Q8MMA1

ALIGNMENTS

RESULT 1

Q9U6Y8 PRELIMINARY; PRT; 225 AA.
AC Q9U6Y8
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DB Fluorescent protein Fp583.
OS Discosoma sp.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=86600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.P., Labas Y.A., Savitsky A.P., Zarskiy A.G.,
RA Markelov M.L., Lukanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Matz M.V., Fradkov A.P., Labas Y.A., Savitsky A.P., Zarskiy A.G.,
RA Markelov M.L., Lukanov S.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF168419; AAF03369.1; -
DR PDB; 1G7K; X-ray; A/B/C/D=1-225.
DR PDB; 1GGX; X-ray; A/B/C/D=1-225.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GLUORESCENT.
DR ProDom; PD01375; Green_fl_protein; 1.
SQ SEQUENCE 225 AA; 25931 MW; FBF9A5369778F689 CRC64;

Alignment Scores:
Pred. No.: 1.59e-108 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 2 Gaps: 0

US-10-006-922A-11 (1-678) x Q9U6Y8 (1-225)

Qy 1 ATGAGGCTCTCCAAAGAAATGTTATCAAGGAGTTCATGAGGTTAAAGGTTTCGCATCGAAGGA 60
Db |||
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGAGAGAGGAGGCCATACGAAGGC 120
Db |||
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTCCTGGGATATT 180
Db |||
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCGACATACCA 240
Db |||
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAAGCTGTCATTTCTCGAAGGATTTAAATGGGAAAGGGTCATGAATTTGAA 300
Db |||
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db |||
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTCAAGTTCATTTGGCGTGAACCTTCTCCGATGACCTGTATGTCMAAAGAAGACA 420
Db |||
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGCGTGGGAAGCCAGCAGCTAGCGTTGTATCTCGTATGCGTGTGAAAGGAGAG 480
Db |||
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAGGCTCTGAAGCTCAAGACCGTGGTCAATTCATCTAGTTCGAATTCAAAGTATT 540
Db |||
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAAGCCCTGTGAGCTACCGGGTACTACTATGTTGATCTCCAAACTGGAT 600
Db |||
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACAGCCCAACAGAGATATACAAATCGTTGAGCAGTATGAAGAACCAGAGGAGCG 660
Db |||
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCCCTT 675
Db |||
Qy 221 HisHisLeuPheLeu 225

RESULT 2

Q66ND7 PRELIMINARY; PRT; 236 AA.

AC Q66ND7
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Enhanced red fluorescent protein R+.
OS Discosoma sp. RC-2004.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=289055;
RN [1]
RP SEQUENCE FROM N.A.
RA Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT "Cloning of Anthozoan Fluorescent Protein Genes."
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY679107; AAU04444.1; -.
DR InterPro; IPR009017; GFP_like.

DR InterPro; IPR011584; GFP related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP; 1.
DR PRODOM; PD013756; Green Fl protein; 1.
SQ SEQUENCE 236 AA; 27032 MW; BB3844BB06829EF0 CRC64;

Alignment Scores:
Pred. No.: 8 83e-107 Length: 236
Score: 1196.00 Matches: 221
Percent Similarity: 99.11% Conservativity: 2
Best Local Similarity: 98.22% Mismatches: 2
Query Match: 98.44% Indels: 0
DB: 2 Gaps: 0

US-10-006-922A-11 (1-678) x Q66ND7 (1-236)

Qy 1 ATGAGGCTCTCCAAAGAAATGTTATCAAGGAGTTCATGAGGTTTAAAGTTCGCATCGAAGGA 60
Db |||
Qy 1 MetSerCysSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGAAGAGAGAGGAGGCCATACGAAGGC 120
Db |||
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTCCTGGGATATT 180
Db |||
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCGACATACCA 240
Db |||
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAAGCTGTCATTTCTCGAAGGATTTAAATGGGAAAGGGTCATGAATTTGAA 300
Db |||
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db |||
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTCAAGTTCATTTGGCGTGAACCTTCTCCGATGAGCTCTGTTATGCAAAAGAAGACA 420
Db |||
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGCGTGGGAAGCCAGCAGCTAGCGTTGTATCTCGTATGCGTGTGAAAGGAGAG 480
Db |||
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAGGCTCTGAAGCTCAAGACCGTGGTCAATTCATCTAGTTCGAATTCAAAGTATT 540
Db |||
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysThrIle 180
Qy 541 TACATGGCAAGAAGCCCTGTGAGCTACCGAGTACTACTATGTTGATCTCCAAACTGGAT 600
Db |||
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACAGCCCAACAGAGTATACAAATCGTTGAGCAGTATGAAGAACCAGAGGAGCG 660
Db |||
Qy 201 IleThrSerHisAsnLysAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCCCTT 675
Db |||
Qy 221 HisHisLeuPheLeu 225

RESULT 3

Q6KF85 PRELIMINARY; PRT; 225 AA.

AC Q6KF85
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Orange fluorescent protein FP586.


```
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysThrIle 180
Qy 541 TACATGGCAAGAGCCTGTCAGCTACCAGGTACTACTATGTTGACTCAAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACCAAGCCACACGAGACTATACAACTCTTCAGCAGATATGAAGAACCGAGGACGC 660
Db 201 IleThrSerHisAnLysAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 5
Q9GTJ7 ID Q9GTJ7 PRELIMINARY; PRT; 230 AA.
AC Q9GTJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Red fluorescent protein.
GN Name=FP593;
OS Discosoma sp. SSAL-2000.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=137428;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2043459; PubMed=10981720; DOI=10.1016/S0014-5793(00)01895-0;
RA Fradkov A.F., Chen Y., Ding L., Barsova E.V., Matz M.V.,
RA Lukyanov S.A.;
RT "Novel fluorescent protein from Discosoma coral and its mutants
RT possesses a unique far-red fluorescence.";
RL FEBS Lett. 479:127-130(2000).
DR EMBL; AF272711; AAG16224.1; -.
DR HSP; Q9U6Y8; 1GX.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
SQ SEQUENCE 230 AA; 26370 MW; 5215B1B436D67E51 CRC64;

Alignment Scores:
Pred. No.: 4.32e-96 Length: 230
Score: 1085.50 Matches: 199
Percent Similarity: 95.58% Conservative: 17
Best Local Similarity: 88.05% Mismatches: 9
Query Match: 89.34% Indels: 1
DB: 2 Gaps: 1

US-10-006-922A-11 (1-678) x Q9GTJ7 (1-230)

Qy 1 ATGAGGTCCTCCAGAAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60
Db 1 MetSerCysSerLysAsnValIleLysGluPheMetArgPheLysValArgMcGluGly 20
Qy 61 ACGGTCAATGGGCACGAGTTTGAATAAGAGCGAAGGAGGAGGAGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleLysGlyGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAAATACCGTAAGCTTAAGTACCAAGGGGACCTTGGCATTTGCTGGGATTT 180
Db 41 HisCysSerValLysLeuMetValThrLysGlyGlyProLeuProPheAlaPheAspIle 60
Qy 181 TTGTCCACCAATTTCACTATGGAAGCAAGGTATATGTCAGACACCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATATAAAGCTGTCTATTTCTCTGAAGGATTTAAATGGGAAGGGTTCATGAACCTTTGAA 300
```

```
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGGCTGCTTACTGTAAACCAGGATTCAGTTTCAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValSerGlnAspSerSerLeuLysAspGlyCysPheIleTyr 120
Qy 361 AAGCTCAAGTTTCAATTCGGGTGAACCTTTCTTCGATGGACCTGTTATGCAAAAAGAGACA 420
Db 121 GluValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnArgArgThr 140
Qy 421 ATGGGCTGGGAAGCCAGCACTAGCGTTTGTATCTCTCGTATGGCTGTTGAAAGGAGAG 480
Db 141 ArgGlyTrpGluAlaSerSerGluArgLeuTyrProArgAspGlyValLeuLysGlyAsp 160
Qy 481 ATTCAATAAGGCTCTGAAGCTGAAAGACGGTGGTTCATTACTAGTTCAATTCAAAGTATT 540
Db 161 IleHisMetAlaLeuArgLeuGlyGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAGAGCT--GTGCAGCTACACAGGCTACTACTATGTTGACTCCAAACTG 597
Db 181 TyrMetValLysLysProSerValGlnLeuProGlyTyrTyrValAspSerLysLeu 200
Qy 598 GATATAACAAGCCACCAAGCACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGA 657
Db 201 AspMetThrSerHisAsnGluAspTyrThrValValGluGlnTyrGluLysThrGlnGly 220
Qy 658 CGCCACCATCTGTTCTT 675
Db 221 ArgHisHisProPheIle 226

RESULT 6
Q66PV1 ID Q66PV1 PRELIMINARY; PRT; 221 AA.
AC Q66PV1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Chromoprotein.
OS Acropora tenuis (Purple tipped acropora).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoenina; Acroporidae; Acropora.
OX NCBI_TaxID=70783;
RN [1]
RP SEQUENCE FROM N.A.
RA Alieva N.O., Melekhvitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646074; AAU06853.1; -.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 221 AA; 25103 MW; 3BEA8B848CF5DA4 CRC64;

Alignment Scores:
Pred. No.: 7.39e-69 Length: 221
Score: 804.00 Matches: 148
Percent Similarity: 79.36% Conservative: 25
Best Local Similarity: 67.89% Mismatches: 45
Query Match: 66.17% Indels: 0
DB: 2 Gaps: 0

US-10-006-922A-11 (1-678) x Q66PV1 (1-221)

Qy 16 AATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGAACGGTCAATGGCAC 75
Db 2 SerValIleAlaLysGlnMetThrTyrLysValTyrMetSerGlyThrValAsnGlyHis 21
Qy 76 GAGTTGAAATAGAACGAGGAGGAGGAGGCCATACGAAGGCCCAATACCGTAAAG 135
```

```

Db      22  TyrPheGluValGluGlyAspGlyLysGlyLysProTyrGluGlyGluInThrValLys 41
Qy      136  CTTAAGTAACCAAGGGGACCTTTGCCATTTGCTTGGGATATTTGTCCACCAATTT 195
Db      42  LeuThrValThrLysGlyProLeuProPheAlaTrpAspIleLeuSerProGlnPhe 61
Qy      196  CAGTATGGAAGCAAGGTATATGTCACAGCACCTGCCACATACAGACTATAAAGCTG 255
Db      62  GlnTyrGlySerIleProPheThrLysTyrProGluAspIleProAspTyrValLysGln 81
Qy      256  TCATTTCTCCTGAAGGATTAAATGGGAAGGGTTCATGAACCTTTGAAGACGGTGGCGTCTG 315
Db      82  SerPheProGluGlyTyrThrTrpGluArgIleMetAsnPheGluAspAlaValCys 101
Qy      316  ACTGTAAACCAAGGATTCAGTTCAGAGTGGCTTTTCCATCTACAAAGGTCAAGTTCATT 375
Db      102  ThrValSerAsnAspSerSerIleGlnGlyAsnGlyMetLeuIleGlyAsnPheMetAlaLeu 161
Qy      496  AAGCTGAAGACGGTGTCTATACCTAGTTGAATTCAAAGTATTTACATGGCAAGAAC 555
Db      162  LysLeuGluGlyGlyHisTyrLeuCysGluPheLysSerThrTyrLysAlaLysLys 181
Qy      556  CTTGTGACGTACACAGGTTACTACTATGTTGACTCCAACTGGATATAAAGCCCAAC 615
Db      182  ProValArgMetProGlyTyrHisTyrValAspArgLysLeuAspValThrAsnHisAsn 201
Qy      616  GAAGACTATACATCGTTGAGCAGTATGAAGAAGACGAGGAGCCACCATCTG 669
Db      202  LysAspTyrThrSerValGluGlnCysGluIleSerIleAlaArgLysProLeu 219

RESULT 7
Q66PUB ID Q66PUB PRELIMINARY; PRT; 221 AA.
AC Q66PUB;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Chromoprotein.
OS Acropora aculeus.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoenina; Acroporidae; Acropora.
OX NCBI_TaxID=287157;
RN [1]
RP SEQUENCE FROM N.A.
RA Aliava N.O., Meleshkevitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646077; AAU06856.1; -.
DR InterPro; IPR011584; GFP_related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 221 AA; 24976 MW; 307DB4FC4B018B43 CRC64;

Alignment Scores:
Pred. No.: 2,81e-68 Length: 221
Score: 798.00 Matches: 145
Percent Similarity: 80.48% Conservative: 24
Best Local Similarity: 69.05% Mismatches: 41
Query Match: 65.68% Indels: 0
DB: 2 Gaps: 0

US-10-006-922a-11 (1-678) x Q66PUB (1-221)

```

```

Qy      16  AATGTTATCAAGAGTTCATGAGTTTAAAGTTTCGCATGGAAGACGGTCAATGGCAC 75
Db      2  SerValleAlaLysGlnMetThrTyrLysValTyrMetSerGlyThrValAsnGlyHis 21
Qy      76  GAGTTTGAATAGCAAGCGAAGGAGGAGGAGCCATACGAAGCCACCAATACCGTAAAG 135
Db      22  TyrPheGluValGluGlyAspGlyLysGlyLysProTyrGluGlyGluInThrValLys 41
Qy      136  CTTAAGTAAACCAAGGGGACCTTTGCCATTTGCTTGGGATATTTGTCCACCAATTT 195
Db      42  LeuThrValThrLysGlyProLeuProPheAlaTrpAspIleLeuSerProGlnSer 61
Qy      196  CAGTATGGAAGCAAGGTATATGTCACAGCACCTGCCACATACAGACTATAAAGCTG 255
Db      62  GlnTyrGlySerIleProPheThrLysTyrProAspAspIleProAspTyrValLysGln 81
Qy      256  TCATTTCTCCTGAAGGATTAAATGGGAAGGGTTCATGAACCTTTCAAGACGGTGGCGTCTG 315
Db      82  SerPheProGluGlyTyrThrTrpGluArgIleMetAsnPheGluAspGlyAlaValCys 101
Qy      316  ACTGTAAACCAAGGATTCAGTTCAGAGTGGCTTTTCCATCTACAAAGGTCAAGTTCATT 375
Db      102  ThrValSerAsnAspSerSerIleGlnGlyAsnGlyPheIleTyrAsnValLysPheSer 121
Qy      376  GGGCTGAACCTTTCTCCTGATGACCTGTTATTCAAAGAACAAATGGGCTGGGAAGCC 435
Db      122  GlyLeuAsnPheProAsnGlyProValMetGlnLysLysThrGlnGlyTrpGluPro 141
Qy      436  AGCACTGAGCGTTGTATCTCGTGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 495
Db      142  AsnThrGluArgLeuPheAlaArgAspGlyMetLeuIleGlyAsnPheMetAlaLeu 161
Qy      496  AAGCTGAAGACGGTGTCTATACCTAGTTGAATTCAAAGTATTTACATGGCAAGAAC 555
Db      162  LysLeuGluGlyGlyHisTyrLeuCysGluPheLysSerThrTyrLysAlaLysLys 181
Qy      556  CTTGTGACGTACACAGGTTACTACTATGTTGACTCCAACTGGATATAAAGCCCAAC 615
Db      182  ProValArgMetProGlyTyrHisTyrValAspArgLysLeuAspValThrAsnHisAsn 201
Qy      616  GAAGACTATACATCGTTGAGCAGTATGAAGAAGACGAGGAGGAGGAGGAGGAGGAGG 645
Db      202  ArgAspTyrThrSerValGluGlnCysGlu 211

RESULT 8
Q66PUB ID Q66PUB PRELIMINARY; PRT; 221 AA.
AC Q66PUB;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Chromoprotein.
OS Acropora millepora (Coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Astrocoenina; Acroporidae; Acropora.
OX NCBI_TaxID=45264;
RN [1]
RP SEQUENCE FROM N.A.
RA Aliava N.O., Meleshkevitch E.A., Field S.F., Matz M.V.;
RT "Survey of coral GFP-like proteins.";
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY646075; AAU06854.1; -.
DR InterPro; IPR009017; GFP_like.
DR InterPro; IPR011584; GFP_related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP.
DR ProDom; PD013756; Green fl protein; 1.
SQ SEQUENCE 221 AA; 24987 MW; 546D2CFPAF2552F2 CRC64;

Alignment Scores:
Pred. No.: 3,51e-68 Length: 221

```



```
Qy 559 GTGCAAGCTACCGGGTACTACTATGTGACTCCAACTGGATATAACAAGCCACACGAA 618
Db 181 ValysMetProGlyTyrHisTyrValAspArgLysLeuAspValThrAsnHisAsnLys 200
Qy 619 GACTATACAAATCGTTGAGCATATGAA 645
Db 201 AspTyrThrSerValGluGlnCysGlu 209

RESULT 12
Q66ND6 PRELIMINARY; PRT; 227 AA.
AC Q66ND6;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Cyan fluorescent protein Cl.
OS Discosoma striata.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=105400;
RN [1]
RP SEQUENCE FROM N.A.
RA Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT "Cloning of Chidarian Fluorescent Protein Genes.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY679108; AAU04445.1; -.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR InterPro; IPR000169; Pept_cys_acsite.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_UORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
DR PROSITE; PS00639; THIOI_PROTEASE_HIS; UNKNOWN_1.
SQ SEQUENCE 227 AA; 25640 MW; 568639306F44B244 CRC64;

Alignment Scores:
Pred. No.: 3.8e-63 Length: 227
Score: 745.00 Matches: 135
Percent Similarity: 75.68% Conservative: 33
Best Local Similarity: 60.81% Mismatches: 54
Query Match: 61.32% Indels: 0
DB: 2 Gaps: 0

US-10-006-922A-11 (1-678) x Q66ND6 (1-227)
Qy 1 ATGAGGCTTCCAGAAATGTTATCAAGAGTTCAAGGTTTAAGGTTTCGATCGAAGGA 60
Db 1 MetSerTrpSerLysSerValIleLysGluMetLeuLeuAsnLeuHisLeuGluGly 20
Qy 61 ACGGTCAATGGCCAGCGAGTTTGAATAGACGCGAAGGAGAGGGGAGCCATACGAAGGC 120
Db 21 ThrPheAsnGlyHisTyrPheGluLeuLysGlyLysGlyLysGlyAsnProAsnGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGATATT 180
Db 41 ThrAsnThrValThrLeuGluValThrLysGlyGlyProLeuProPheGlyTrpHisLe 60
Qy 181 TTGTCACCAACAATTCAGTATGGAAGCAAGATATATGTCAGACCCCTGCCGACATACCA 240
Db 61 LeuCysProGlnPheGlnTyrGlyAsnLysAlaPheValHisHisProAspAspIlePro 80
Qy 241 GACTATAAAAGCTGCTATTTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db 81 AspTyrLeuLysLeuSerPheProGluGlyTyrThrTrpGluArgSerMetHisPheGlu 100
Qy 301 GACGGTGGCGTCTGTTACTACTAACCCAGGATTCAGTTTTCAGAGTGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyLeuCysCysIleThrAsnAspIleSerLeuThrGlyAsnCysPheAsnTyr 120
Qy 361 AAGTCAAGTTTCATGGGTGAACCTTCTCCGATGAGACCTGTATGCAAAAAGACACA 420
Db 121 AspIleLysPheThrGlyLeuAsnPheProAsnGlyProValValGlnLysLysThr 140
```

```
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATCGCGTGTGAAGGAGAG 480
Db 141 ThrGlyTrpGluProSerThrGluArgLeuTyrProArgAspGlyValLeuIleGlyAsp 160
Qy 481 ATTCAATAGGCTCTGAAGCTGAAGACGGTGGTTCATTACCTAGTGTGAATTCAAAAGTATT 540
Db 161 IleHisHisAlaLeuThrValAlaGlyGlyGlyHisTyrValCysAspIleLysThrVal 180
Qy 541 TACATGGCAAGAAGCGCTGTGCAGCTACCAAGGCTACTACTATGTTGACTCCAACTGGAT 600
Db 181 TyrArgAlaLys**ProValLysMetProGlyTyrHisTyrValAspProLysLeuVal 200
Qy 601 ATAAACAGCCACAAAGACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGGACGC 660
Db 201 IleArgSerAsnAspLysGluPheMetLysValGluGluHisGluLeuAlaValAlaArg 220
Qy 661 CACCAT 666
Db 221 ***His 222

RESULT 13
Q906Y7 PRELIMINARY; PRT; 232 AA.
AC Q906Y7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Fluorescent protein FP483.
OS Discosoma striata.
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Corallimorpharia;
OC Discosomatidae; Discosoma.
OX NCBI_TaxID=105400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99436614; PubMed=10504696;
RA Matz M.V., Fradkov A.F., Labas Y.A., Savitsky A.P., Zaraisky A.G.,
RA Markelov M.L., Lukyanov S.A.;
RT "Fluorescent proteins from nonbioluminescent Anthozoa species.";
RL Nat. Biotechnol. 17:969-973 (1999).
DR EMBL; AF168420; AAF03370.1; -.
DR HSSP; Q906Y8; IGGX.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP-related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_UORESCENT.
DR ProDom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 232 AA; 26435 MW; AA8F18EE283CB4D CRC64;

Alignment Scores:
Pred. No.: 1.2e-61 Length: 232
Score: 729.50 Matches: 133
Percent Similarity: 76.00% Conservative: 38
Best Local Similarity: 59.11% Mismatches: 53
Query Match: 60.04% Indels: 1
DB: 2 Gaps: 1

US-10-006-922A-11 (1-678) x Q906Y7 (1-232)
Qy 1 ATGAGGCTTCCAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGATCGAAGGA 60
Db 1 MetSerCysSerLysSerValIleLysGluMetLeuLeuAspLeuHisLeuGluGly 20
Qy 61 ACGGTCAATGGCCAGCGAGTTTGAATAGACGCGAAGGAGAGGGGAGCCATACGAAGGC 120
Db 21 ThrPheAsnGlyHisTyrPheGluLeuLysGlyLysGlyLysGlyGlnProAsnGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGGTTAAACCAAGGGGGACCTTTGCCATTTGCTTGGATATT 180
Db 41 ThrAsnThrValThrLeuGluValThrLysGlyLysGlyLysGlyLysGlyTrpHisLe 60
```

```

QY 181 TTGTCCACCAATTTCTAGTATGGAAGCAAGTATATCTCAAGCACCTGCGGCATACCA 240
Db 61 LeuCysProGlnPheGlnTyrGlyAsnLysAlaPheValHisHisProAspAsnIleHis 80
QY 241 GACTATATAAAGCTGTCTATTTCTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTTGAA 300
Db 81 AspTyrLeuLysLeuSerPheProGluGlyTyrThrTrpGluArgSerMetHisPheGlu 100
QY 301 GACGGTGGCGTGTCTACTGTAAACCCAGGATTCAGTTTTCAGGAGTGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyLeuCysCysIleThrAsnAspIleSerLeuThrGlyAsnCysPheTyrTyr 120
QY 361 AAGGTCAAGTTCATTCGCGGTGAACCTTCTCCGATCGACTGTATGCAAAAGAGACA 420
Db 121 AspIleLysPheThrGlyLeuAsnPheProAsnGlyProValValGlnLysLysThr 140
QY 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTCGTATGGCGTGTGAAAGGAGAG 480
Db 141 ThrGlyTrpGluProSerThrGluArgLeuTyrProArgaspGlyValLeuIleGlyAsp 160
QY 481 ATTCAATAGGCTCGAAGCTGAAGAGCGTGGTTCATTCAGTGTGAATTCAAAGTATT 540
Db 161 IleHisHisAlaLeuThrValGluGlyGlyGlyHisTyrAlaCysAspIleLysThrVal 180
QY 541 TACATGCAAGAG---CTGTGACACTACCAGGCTACTACTACTGTGCTCCAAACTG 597
Db 181 TyrArgAlaLysLysAlaLeuLysMetProGlyTyrHisTyrValAspThrLysLeu 200
QY 598 GATATAACAAGCCACAAGAGACTATACAATCTCTGAGCAGTATGAAGAACCGAGGA 657
Db 201 ValIleTrpAsnAspLysGluPheMetLysValGluGluHisGluIleAlaValAla 220
QY 658 CGCCACCATCTGTTTC 672
Db 221 ArgHisProPhe 225

RESULT 14
Q963F5 PRELIMINARY; PRT; 225 AA.
ID Q963F5 AC Q96ND3
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Green fluorescent protein.
OS Montastrea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastrea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
RA Falkowski P., Gorbunov M., Kolber Z.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF384683; AAK62982.2;
DR HSP; Q906Y8; ICGX.
DR GO; GO:0006091; P:energy pathways; IEA.
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP related.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_L1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 225 AA; 25947 MW; 77DE7D7C616929AF CRC64;

Alignment Scores:
Pred. No.: 7,95e-58 Length: 225
Score: 690.00 Matches: 120
Percent Similarity: 75.23% Conservative: 44
Best Local Similarity: 55.05% Mismatches: 54
Query Match: 56.79% Indels: 0
DB: 2 Gaps: 0

```

US-10-006-922A-11 (1-678) x Q963F5 (1-225)

```

QY 16 AATGTTATCAAGGTTTCATGAGGTTTAAGGTTTCGCGATCGAAGGACCGTCAATGGGCAC 75
Db 2 SerValIleLysProIleMetGluIleLysLeuArgMetGlnGlyValAlaAsnGlyHis 21
QY 76 GAGTTTGAATAGAACGGGAAGGAGGAGGAGGCATACGAAGGCCCAATACCGTAAAG 135
Db 22 LysPheValIleLysGlyGluGlyGlyProPheGluGlyThrGlnThrIleAsn 41
QY 136 CTTAAGGTAAACCAAGGGGACCTTTGCGCATTTGCTTGGGATATTTTGTCCACCAATTT 195
Db 42 LeuThrValLysGluGlyAlaProLeuProPheAlaTyrAspIleLeuThrSerAlaPhe 61
QY 196 CAGTATGGAAGCAAGTATATGTCACGACCCCTCCGACATACACAGACTATAAAGCTG 255
Db 62 GlnTyrGlyAsnArgValPheThrLysTyrProAspAspIleProAspTyrPheLysGln 81
QY 256 TCATTTCCTCAAGGATTTAAATGGGAAGGCTCATGAACCTTTGAAGACGGTGGCGTCTT 315
Db 82 ThrPheProGluGlyTyrSerTrpGluArgIleMetAlaTyrGluAspGlnSerIleCys 101
QY 316 ACTGTAAACCCAGGATTCAGTTTTCAGGATGGCTGTTTTCATCTACAAAGTCAAGTTCATT 375
Db 102 ThrAlaThrSerAspIleLysMetGluGlyAspCysPheIleTyrGluIleGlnPheHis 121
QY 376 GCGGTGACACTTTCTCCGATGACCTGTTATGCANAAGAGACAATGGGCTGGGAAGCC 435
Db 122 GlyValAsnPheProAsnGlyProValMetGlnLysLysThrLeuLysThrGluPro 141
QY 436 AGCACTGAGGCTTTGTTATCTCTGATGGCGTGTGAAAGAGAGAGATTTCATAGGCTCTG 495
Db 142 SerThrGluLysMetTyrValArgAspGlyValLeuLysGlyAspValAsnMetAlaLeu 161
QY 496 AAGCTGAAAGACCGTGTCTATTACCTAGTTGAAATTCAAAGTATTTACATGGCAAGAG 555
Db 162 LeuLeuGluGlyGlyGlyHisTyrArgCysAspPheArgSerThrTyrLysAlaLysLys 181
QY 556 CTTGTGCGAGCTACAGGTTACTACTATGTTGACTCCAAACTGGATATACAGGCCAAC 615
Db 182 ArgValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeuSerHisAsp 201
QY 616 GAAGACTATCAATCGTTGAGCAGTATGAAGAAGACCGAGGAGCGCCACCATCTG 669
Db 202 AsnAspTyrAsnThrValLysLeuSerGluAspAlaGluAlaArgTyrSerMet 219

RESULT 15
Q66ND3 PRELIMINARY; PRT; 225 AA.
ID Q66ND3 AC Q66ND3
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Green fluorescent protein G2.
OS Montastrea cavernosa (great star coral).
OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Scleractinia;
OC Faviina; Faviidae; Montastrea.
OX NCBI_TaxID=63558;
RN [1]
RP SEQUENCE FROM N.A.
RA Carter R.W., Gibbs P.D.L., Schmale M.C.;
RT "Cloning of Cnidarian Fluorescent Protein Genes.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY679111; AAU04448.1;
DR InterPro; IPR009017; GFP like.
DR InterPro; IPR011584; GFP related.
DR InterPro; IPR000786; Green_fl_protein.
DR Pfam; PF01353; GFP; 1.
DR PRINTS; PR01229; GFP_L1.
DR PRODOM; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 225 AA; 25910 MW; A744E8D6A6C423AE CRC64;

Alignment Scores:
Pred. No.: 3.02e-57 Length: 225
Score: 684.00 Matches: 120

```

Percent Similarity: 73.85% Conservative: 41
Best Local Similarity: 55.05% Mismatches: 57
Query Match: 56.30% Indels: 0
DB: 2 Gaps: 0

US-10-006-922A-11 (1-678) x Q66ND3 (1-225)

```
QY 16 AATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGAAGGAACGGTCAATGGGCAC 75
   ::::::::::::::::::::
Db 2 ServalleysProaspMetLysIleLysLeuArgMetGluGlyAlaValasnGlyHis 21
   ::::::::::::::::::::

QY 76 GAGTTTGAATAGAACGCGAGGAGGAGGCGGCATACGAAGGCCACAATACCGTAAAG 135
   ::::::::::::::::::::
Db 22 AsnPheValIleGluGlyGluLysGlyLysProPheGluGlyThrGlnThrIleAsn 41
   ::::::::::::::::::::

QY 136 CTTAAGGTATACCAAGGGGACCTTTCGCTTGGGATATTTTGTCCACCATTTT 195
   ::::::::::::::::::::
Db 42 LeuThrValLysGluGlyGlyProLeuProPheAlaTyrAspIleLeuThrAlaAlaPhe 61
   ::::::::::::::::::::

QY 196 CAGTATGGAAGCAAGGTATATGTCAAGCACCTCGCCACATACCAGACTATATAAAGCTG 255
   ::::::::::::::::::::
Db 62 GlnTyrGlyAsnArgAlaPheThrLysTyrProArgAspIleAlaAspTyrPheLysGln 81
   ::::::::::::::::::::

QY 256 TCATTTCTCAAGGATTAAATGGGAAAGGTCATGAACCTTTGAAGACGGTGGCGTCGTT 315
   ::::::::::::::::::::
Db 82 SerPheProGluGlyTyrSerTrpGluArgSerMetThrTyrGluAspGlnGlyIleCys 101
   ::::::::::::::::::::

QY 316 ACTGTAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTACAAGGTCAGTTTCATT 375
   ::::::::::::::::::::
Db 102 IleIleLysSerAspIleArgMetGluGlyAspCysPheIleTyrGluIleArgTyrAsp 121
   ::::::::::::::::::::

QY 376 GCGGTGAACCTTTCCTCCGATGGACCTGTATGCAAAAGAACAAATGGGCTGGGAAGCC 435
   ::::::::::::::::::::
Db 122 GlyValAsnPheProProSerGlyProValMetGlnLysLysThrLeuLysTrpGluPro 141
   ::::::::::::::::::::

QY 436 AGCACTGAGCGTTTGTATCCTCGTGATGGCGTGTGAAAGGAGAGATTTCATAAGGCTCTG 495
   ::::::::::::::::::::
Db 142 SerThrGluLysMetTyrValArgAspGlyValLeuLysGlyGluValAsnMetAlaLeu 161
   ::::::::::::::::::::

QY 496 AAGCTGAAGACCGGTGATCTACCTAGTTCAATTCAAAAGTATTTACATGGCNAAGAG 555
   ::::::::::::::::::::
Db 162 LeuLeuGluGlyGlyGlyHisTyrArgCysAspPheArgSerThrTyrLysAlaLysLys 181
   ::::::::::::::::::::

QY 556 CCTGTGCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGATATAACAAGCCACAAC 615
   ::::::::::::::::::::
Db 182 ArgValGlnLeuProAspTyrHisPheValAspHisArgIleGluIleLeuSerHisAsp 201
   ::::::::::::::::::::

QY 616 GAAGACTATACATCGTTGAGCAGTATGAAGAACCCAGGAGCCGCCACCATCTG 669
   ::::::::::::::::::::
Db 202 AsnAspTyrAsnThrValLysLeuSerGluAsnAlaGluAlaArgTyrSerMet 219
   ::::::::::::::::::::
```

Search completed: June 30, 2005, 16:04:10
Job time : 166.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 30, 2005, 15:46:45 ; Search time 136.5 Seconds
(without alignments)
3842.103 Million cell updates/sec

Title: US-10-006-922a-11

Perfect score: 1215

Sequence: 1 atgagggtcttccaagaatgt.....gccaccatctgttcctttaa 678

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 4211384

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool_P/US1006922/runat 30062005 110551 9404/app query.fasta_1.839
-DB=A_Geneseq -QFMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40 cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptco -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US1006922@cgn 1_1_224 @runat 30062005 110551 9404 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_16Dec04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001s.*
5: Geneseq2002s.*
6: Geneseq2003as.*
7: Geneseq2003bs.*
8: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1214	99.9	225	3 AAY99836	Aay99836 Discosoma
2	1214	99.9	225	3 AAB01622	Aab01622 Discosoma
3	1214	99.9	225	4 AAG65509	Agg65509 Anthozoan
4	1214	99.9	225	5 AAB08834	Abb08834 Yeast opt
5	1214	99.9	225	5 AAE28833	Aae28833 Discosoma
6	1214	99.9	225	5 AAE17540	Aae17540 Discosoma
7	1214	99.9	225	5 AAO18270	Aao18270 Discosoma
8	1214	99.9	225	6 AAE34962	Aae34962 Discosoma
9	1214	99.9	225	7 ADC24126	Adc24126 Discosoma
10	1214	99.9	225	7 ABW00918	Abw00918 Discosoma

11	1214	99.9	225	7 ADF70403	Adf70403 Discosoma
12	1214	99.9	225	7 ADH34489	Adh34489 Discosoma
13	1214	99.9	225	7 ADL46203	Adl46203 Discosoma
14	1214	99.9	225	7 ADN33979	Adn33979 Wild-type
15	1214	99.9	225	8 ADI36421	Adi36421 Discosoma
16	1214	99.9	225	8 ADM97769	Adm97769 D sp red
17	1214	99.9	225	8 ADQ28780	Adq28780 Jellyfish
18	1214	99.9	487	5 ABB08821	Abb08821 Autofluor
19	1214	99.9	506	5 ABB08822	Abb08822 Autofluor
20	1214	99.9	547	5 ABB08823	Abb08823 Autofluor
21	1211	99.7	225	5 AAE28920	Aae28920 Discosoma
22	1211	99.7	225	7 ABW00937	Abw00937 Discosoma
23	1211	99.7	225	7 ABW00929	Abw00929 Discosoma
24	1211	99.7	225	7 ABW00938	Abw00938 Discosoma
25	1211	99.7	225	7 ADH34498	Adh34498 Discosoma
26	1211	99.7	225	7 ADH34499	Adh34499 Discosoma
27	1210	99.6	225	5 AAE28919	Aae28919 Discosoma
28	1210	99.6	225	7 ABW00930	Abw00930 Discosoma
29	1210	99.6	226	4 AAG65510	Agg65510 Anthozoan
30	1210	99.6	242	7 ADE24109	Ade24109 Discosoma
31	1210	99.6	545	7 ADL18132	Adl18132 RFP:PS (HI
32	1210	99.6	548	7 ADL18156	Adl18156 RFP:PS (HI
33	1209	99.5	225	5 AAE28922	Aae28922 Discosoma
34	1209	99.5	225	7 ABW00936	Abw00936 Discosoma
35	1209	99.5	225	7 ABW00931	Abw00931 Discosoma
36	1209	99.5	226	5 ABB08835	Abb08835 Yeast opt
37	1208	99.4	225	7 ABW00932	Abw00932 Discosoma
38	1208	99.4	225	7 ABW00935	Abw00935 Discosoma
39	1208	99.4	225	7 ABW00939	Abw00939 Discosoma
40	1208	99.4	225	7 ADH34500	Adh34500 Discosoma
41	1207	99.3	225	5 AAE28921	Aae28921 Discosoma
42	1207	99.3	225	5 AAE17541	Aae17541 Discosoma
43	1207	99.3	225	7 ABW00941	Abw00941 Discosoma
44	1207	99.3	225	7 ABW00940	Abw00940 Discosoma
45	1207	99.3	225	7 ADL46222	Adl46222 Discosoma

ALIGNMENTS

RESULT 1

AAY99836

ID AAY99836 standard; protein; 225 AA.

XX AAY99836;

XX

DT 12-SEP-2003 (revised)

DT 19-SEP-2000 (first entry)

XX

DE Discosoma sp. "red" novel fluorescent protein drFP583.

XX

KW Anthozoa; drFP583; fluorescent protein; non-bioluminescent organism; fluorescent labeling.

KW

OS Discosoma sp; "red".

XX

FH Key Location/Qualifiers

FT Misc-difference 122

FT /note= "encoded by TC"

FT Misc-difference 127

FT /note= "encoded by GTTC"

XX

PN WO200034326-A1.

XX

PD 15-JUN-2000.

XX

PF 10-DEC-1999; 99WO-US029473.

XX

PR 11-DEC-1998; 98US-00210330.

PR

PR 14-OCT-1999; 99US-00418529.

XX

PA (CLON-) CLONTECH LAB INC.

XX

PI Lukyanov SA, Fradkov AF, Labas YA, Matz MV, Green G, Chen Y;

```

PI Ding L;
XX WPI; 2000-423381/36.
XX Novel fluorescent protein from non-bioluminescent Discosoma sp. red,
PT useful for fluorescent labeling and as markers.
XX Claim 20; Page 74-75; 86pp; English.
XX The present sequence is a novel fluorescent protein (nFP) encoded by the
CC full-length cDNA drFP583. drFP583 was isolated from Discosoma sp. "red",
CC a non-bioluminescent species of the Class Anthozoa. Fluorescent proteins
CC can be used in fluorescent labeling, a useful tool for marking a protein,
CC cell or organism of interest. Unlike other markers used in protein
CC labeling, such as beta-galactosidase and luciferase, fluorescent proteins
CC do not require an exogenous cofactor or substrate. Methods involving
CC fluorescent proteins are also less laborious and less difficult to
CC control than the traditional methods of fluorescent labeling, where a
CC protein of interest is purified and then covalently conjugated to a
CC fluorophore derivative. Novel fluorescent proteins isolated from species
CC of the Class Anthozoa can be used as markers for gene expression and
CC protein localization studies, and in fluorescence resonance energy
CC transfer (FRET) reactions. They may have improved properties and better
CC suitability for larger excitations compared to prior art fluorescent
CC proteins such as green fluorescent protein. (Updated on 12-SEP-2003 to
CC standardise OS field)
XX Sequence 225 AA;

Alignment Scores:
Pred. No.: 3,01e-135 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 3 Gaps: 0

US-10-006-922A-11 (1-678) x AAY99836 (1-225)

Qy 1 ATGAGGCTCCAGAAATGTTATCAAGAGTTTCATAGGTTTAAGGTTTCGATCGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

Qy 61 ACGGTCAATGGCGACGAGTTTGAATAAGACGCGAAGGAGAGGGAGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40

Qy 121 CACAAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTTGGGTGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

Qy 181 TTGTCACCAATTTTCAGTATGGAAGCAGAGTATATGTCAAGCACCTCGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

Qy 241 GACTATAAAAAGCTGTCAATTTCTGAGGATTAAATGGGAAAGGGTCATGAACCTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheLeu 100

Qy 301 GACGGTGCAGCGTCTACTGTAAACAGGATTCCAGTTTGACAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Qy 361 AAGTCAAGTTCATGGGTGAACCTTTCTCCATGAGACCTGTATGCAAAAGACACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

Qy 421 ATGGCGTGGGAAGCCAGCACGTAGCGTTTGTATCTCGTGATGGCTGTTCAAAAGGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgPheGlyValLeuLysGlyGlu 160

Qy 481 ATTCATAAGGCTCTGAAGCTGAAGACGGTGGTCAATTCCTAGTTGAATTCAAAGTATT 540
Db 481

```

161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

541 TACATGGCAAGAGCGCTGTGAGCTACACGGGTACTACTATCTGACTCCAAACTGGAT 600

181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

601 ATAACAAGCCACAACGAAGACTATACATCGTTGAGCAGTATCAAAAGAACCCGAGGACGC 660

201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220

661 CACCATCTGTCTCTT 675

221 HisHisLeuPheLeu 225

RESULT 2

AAB01622

ID AAB01622 standard; protein; 225 AA.

XX

AC AAB01622;

XX

DT 12-DEC-2000 (first entry)

XX

XX Discosoma sp. red fluorescent protein drFP583.

DE

XX Anthozoa; fluorescent protein; fluorescence labeling; amFP486; cFP484;

KW zFP506; zFP538; drFP583; dsFP483; asFP600; dgFP512; dmFP592.

KW

XX Discosoma sp.

OS

XX WO2000034526-A1.

PN

XX 15-JUN-2000.

PD

XX 10-DEC-1999; 99WO-US029405.

PF

XX 11-DEC-1998; 98US-00210330.

PR

XX (CLON-) CLONTECH LAB INC.

PA

XX Lukyanov SA, Fradkov AF, Labas YA, Matz MV;

PI

XX WPI; 2000-423451/36.

DR

XX Novel method for identifying a DNA sequence encoding fluorescent proteins from non-bioluminescent Anthozoa which are useful for fluorescent labeling and as markers.

PT

XX Claim 3; Page 68-69; 73pp; English.

PS

XX The present sequence is Discosoma sp. red fluorescent protein drFP583. It was isolated using the Aequoria victoria green fluorescent protein (GFP) sequence, which was used to design PCR primers which might isolate other fluorescent proteins from a number of species of Anthozoa. These were Anemonia majano, Clavularia sp., Zoanthus sp., Discosoma sp. and Anemonia sulcata. The cDNA obtained was then screened in the search for sequences encoding fluorescent proteins. The other proteins found in this manner were cFP484, zFP506, zFP538, amFP486, dsFP483, asFP600, dgFP512 and dmFP592. These proteins can be used as fluorescent labels (for gene expression and protein localisation studies and in fluorescence resonance energy transfer (FRET) studies) in place of fluorophore derivatives and luciferases, as these involve laborious processes and the latter require cofactors. They can also be used in place of GFP, which is too stable to be useful when studying short-term or repetitive events

XX

SQ Sequence 225 AA;

Alignment Scores:

Pred. No.: 3,01e-135 Length: 225

Score: 1214.00 Matches: 225

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 99.92% Indels: 0

DB: 3 Gaps: 0

```

US-10-006-922A-11 (1-678) x AAB01622 (1-225)
QY 1 ATGAGGCTTCCAGAGATGTTATCAAGAGTTCATGAGGTTTAAGGTTCCATGGAAGGA 60
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACGGTCAATGGCCAGGAGTTGAATAGAGCCCAAGGAGGAGGAGCCATCAGAGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGACCTTGGCCATTGCTTGGGATATT 180
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 TTGTACACCAATTTCTAGTATGGAAGCAAGATATATGCAAGCACCTGCCGACATACCA 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTATAAAGCTGTCATTTCTCTGAAGGATTTAAATGGGAAGGTCATGAACCTTTGAA 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGTGGCGTCGTACTCTGAACCCAGGATTCAGTTTGCAGGATGCTGTTTCATCTAC 360
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTCAAGTTCATTTGGCGTGAACTTTCTCTCCGATGGACCTTGTATGCAAAAGAGACA 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGGTGGGAAGCCAGCACTGAGCGTTGTATCTCGTGATCGCGTGTGAAAGGAGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATTCATAGGCTCTGAAGCTGAAGACGCTGCTGCTTACCTAGTGTGAATTCAAAGATTT 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCAAGAGCCTGTGACCTACCGAGTACTACTATGTTGACTCCAACTGGAT 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATACAGCCACACAGAGACTATACATGTTGAGCAGATGAAAGAACCGAGGAGCG 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCATCTGTTCCCTT 675
DB 221 HisHisLeuPheLeu 225

RESULT 3
AAG65509
ID AAG65509 standard; protein; 225 AA.
AC AAG65509;
XX
XX 30-NOV-2001 (first entry)
XX
XX Anthozoan red fluorescent protein sequence.
XX
XX Fluorescent protein; Anthozoan; fluorescence; marker; FRET; red.
XX
XX Anthozoa.
XX
XX WO200162919-A1.
XX
XX 30-AUG-2001.
XX
XX 13-FEB-2001; 2001WO-US004625.
XX
XX 23-FEB-2000; 2000US-0184732P.
XX

```

(AURO-) AURORA BIOSCIENCES CORP.
 Nelson D, Zamaira E, Tsien R;
 WPI; 2001-557704/62.
 Proteins for Fluorescence Resonance Energy Transfer (FRET) comprise functional red fluorescent proteins, and the encoding nucleic acids, with key mutations for improving the proteins function.
 Disclosure; Page 85; 90pp; English.
 The invention provides a nucleic acid encoding functional red fluorescent CC protein (II) that differs from the sequence of an Anthozoan red fluorescent protein by at least one amino acid substitution, and with different fluorescent properties. The red fluorescent protein of the invention can be expressed by standard recombinant methodology. (II) are used a fluorescent markers and FRET partners. It is used for identifying protein-protein interactions. (II) is also suitable for multiplexed fluorescent analysis and FRET-based applications using existing Acqorea fluorescent proteins. (II) has improved brightness, reduced spectral cross talk, and is rapidly and efficiently expressed in mammalian cells. The key mutations in the encoding nucleic acids provide improved folding, brightness, and create (II) with sharper, more defined excitation and emission peaks when expressed in mammalian cells. The present sequence represents an anthozoan fluorescent protein

Sequence 225 AA:
 Alignment Scores:
 Pred. No.: 3 01e-135 Length: 225
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.92% Indels: 0
 DB: 4 Gaps: 0
 US-10-006-922A-11 (1-678) x AAG65509 (1-225)

QY 1 ATGAGGCTTCCAGAGATGTTATCAAGAGTTCATGAGGTTTAAGGTTCCATGGAAGGA 60
 DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACGGTCAATGGCCAGGAGTTGAATAGAGCCCAAGGAGGAGGAGCCATCAGAGGC 120
 DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGACCTTGGCCATTGCTTGGGATATT 180
 DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 TTGTACACCAATTTCTAGTATGGAAGCAAGATATATGCAAGCACCTGCCGACATACCA 240
 DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTATAAAGCTGTCATTTCTCTGAAGGATTTAAATGGGAAGGTCATGAACCTTTGAA 300
 DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGGTGGCGTCGTACTCTGAACCCAGGATTCAGTTTGCAGGATGCTGTTTCATCTAC 360
 DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGGTCAAGTTCATTTGGCGTGAACTTTCTCTCCGATGGACCTTGTATGCAAAAGAGACA 420
 DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGGTGGGAAGCCAGCACTGAGCGTTGTATCTCGTGATCGCGTGTGAAAGGAGAG 480
 DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATTCATAGGCTCTGAAGCTGAAGACGCTGCTGCTTACCTAGTGTGAATTCAAAGATTT 540
 DB 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGGCAAGAGCCTGTGACCTACCGAGTACTACTATGTTGACTCCAACTGGAT 600
 DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATACAGCCACACAGAGACTATACATGTTGAGCAGATGAAAGAACCGAGGAGCG 660
 DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCATCTGTTCCCTT 675
 DB 221 HisHisLeuPheLeu 225

Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGGCAAGAGCCTGTCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATAAACAAGCCACAACGAAGACTATACAATCGTTCAGCAGAGTATGAAGAACCAGGAGCGC 660
 Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCATCTGTTCTT 675
 Db 221 HisHisLeuPheLeu 225

RESULT 4

ABB08834
 ID ABB08834 standard; protein; 225 AA.

AC ABB08834;

DT 29-MAY-2002 (first entry)

XX Yeast optimised RFP SEQ ID NO 17.

DE Yeast; red fluorescent protein; RFP; plant; transgenic; GFP;
 KW Saccharomyces cerevisiae; Nicotiana tabacum; Arabidopsis thaliana;
 KW Escherichia coli; green fluorescent protein; biotechnology.

XX Anthozoa.

XX DE20001395-UI.

XX 15-MAR-2001.

XX 27-JAN-2000; 2000DE-02001395.

XX 27-JAN-2000; 2000DE-02001395.

XX (GPCB-) GPC BIOTECH AG.

XX WPI; 2002-228394/29.

DR N-PSDB; ABA95905, ABA95921, ABA95922.

XX New DNA encoding red fluorescent protein, useful as marker in
 PT biotechnology, has sequence optimized for expression in eukaryotes,
 PT especially yeast or plants.

XX Disclosure; Page 13-14; 19pp; German.

XX The invention relates to DNA (I) containing either sequence ABA95905 or
 CC sequence ABA95906 encoding a yeast optimised Red Fluorescent Protein
 CC (yRFP). (I) are used to express red fluorescent protein (RFP) in
 CC eukaryotes, especially yeast, especially Saccharomyces cerevisiae and
 CC plants, especially dicotyledonous plants including Nicotiana tabacum or
 CC Arabidopsis thaliana and also in prokaryotes, especially bacteria,
 CC especially Escherichia coli. RFP is useful in the same way as green
 CC fluorescent protein but is more generally applicable in modern
 CC biotechnology. (I) are optimised for expression in yeast and so generate
 CC RFP at higher levels with stronger fluorescence and thus lowers the
 CC detection limit and gives a better signal-to-noise ratio. The present
 CC sequence is that of the yeast optimised RFP

XX Sequence 225 AA;

Alignment Scores:
 Pred. NO.: 3.01e-135 Length: 225
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.92% Indels: 0
 DB: 5 Gaps: 0

US-10-006-922A-11 (1-678) x ABB08834 (1-225)

QY 1 ATGAGGTCTTCCAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAAGGTTTCGCATGGAAGGA 60
 Db 1 MecArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 QY 61 ACGGTCAATGGGCACGAGTTTGAATAGAGCGAAGGAGGAGGAGGCCATACGAAGGC 120
 Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 QY 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTCCTTCGGGATATT 180
 Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 QY 181 TTGTCAACCAATTCAGTATGGAACCAAGGTATATGTCAAGCACCTGCCGACATACCA 240
 Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 QY 241 GACTATAAAAAGCTGTCTTCTCTGAAGGATTTAAATGGGAAAGGTCATGAATTTGAA 300
 Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 QY 301 GACGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTGAGGATGGCTGTTTCATCTAC 360
 Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 QY 361 AAGGTCAAAGTTCAATTGGCGTGAACCTTTCCTTCCGATGGACCTGTTATGCAAAAAGAGACA 420
 Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 QY 421 ATGGCGTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATCGCGTGTGAAAGGAGAG 480
 Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 QY 481 ATTCAATAGGCTCTGAAGCTGAAAGAGGTTGTCATTACCTAGTGAATTCAAAGTATT 540
 Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
 QY 541 TACATGGCAAGAGCCTGTCAGCTACCAGGGTACTACTATGTTGACTCCAAACTGGAT 600
 Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
 QY 601 ATAAACAAGCCACAACGAAGACTATACAATCGTTCAGCAGAGTATGAAGAACCAGGAGCGC 660
 Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 QY 661 CACCATCTGTTCTT 675
 Db 221 HisHisLeuPheLeu 225

RESULT 5
 AAE28833
 ID AAE28833 standard; protein; 225 AA.
 XX AAE28833;
 XX 27-DEC-2002 (first entry)
 DT Discosoma sp. drFP583 (NFP-6) wild-type protein.
 DE Fluorescent protein; chromoprotein; protease cleavage assay; filter;
 KW fluorescence activated cell sorting application; fluorescent timer;
 KW biosensor; fluorescence resonance energy transfer application; FRET;
 KW colouring agent; recombinant DNA application; analyte detection assay;
 KW sunscreen; second messenger detector; drFP583 protein; NFP-6.
 XX Discosoma sp.
 XX WO200268459-A2.
 XX 06-SEP-2002.
 XX 20-FEB-2002; 2002WO-US005749.


```
PR 21-FEB-2001; 2001US-0270983P.
XX 04-DEC-2001; 2001US-00006922.
PA (CLON-) CLONTECH LAB INC.
XX Lukyanov S, Lukyanov K, Yanushevich Y, Savitsky A, Fradkov A;
XX WPI; 2002-691654/74.
XX N-PSDB; AAD46278.
XX New nucleic acid encoding a non-aggregating chromo- or fluorescent mutant
XX of an aggregating Cnidarian chromo- or fluorescent protein or mutant for
XX analyte detection assays or fluorescence activated cell sorting
XX applications.
XX Disclosure; Page 70-71; 80pp; English.
XX The invention relates to nucleic acid molecules encoding non-aggregating
XX chromo/fluorescent proteins and their mutants. Chromo/fluoro proteins are
XX useful in analyte detection assays, as colouring agents, as markers in
XX recombinant DNA applications, as sunscreens or filters, in fluorescence
XX resonance energy transfer (FRET) applications, as biosensors in
XX prokaryotic and eukaryotic cells, in screening assays, as second
XX messenger detectors, in fluorescence activated cell sorting applications,
XX in protease cleavage assays or as fluorescent timers. The present
XX invention is Discosoma sp. drFP583 (NFP-6) wild-type protein of the
XX sequence 225 AA;
XX SQ Sequence 225 AA;
XX Alignment Scores:
XX Pred. No.: 3.01e-135 Length: 225
XX Score: 1214.00 Matches: 225
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 99.92% Indels: 0
XX DB: 5 Gaps: 0
XX
XX US-10-006-922A-11 (1-678) x AAE28333 (1-225)
XX
XX QY 1 ATGAGGTCCTCCAGAGTATGTTATCAAGGAGTTTCATGAGTTTAAAGTTTCATGGAAGGA 60
XX DB 1 MetArgSerSerLysAsnValIleGlyGluPheMetArgPheLysValArgMetGluGly 20
XX
XX QY 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGGAGGAGGAGGAGGCCATACGAAAGGC 120
XX DB 21 ThrValAsnGlyHisGluPheGluIleGlyGlyGluGlyArgProTyrGluGly 40
XX
XX QY 121 CACATACCGTAAGCTTAAGGTAACCAAGGGGGACCTTGGCATTTGCTTGGGATATT 180
XX DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
XX
XX QY 181 TTGTCACCAATTTTCATGATGGAAGCAAGGTATATGTCAGGACCTGCGGCATACCA 240
XX DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
XX
XX QY 241 GACTATATAAAGCTGTCTATTTCTTGAAGGATTTAAATGGAAAGGGTCATCAACTTTGAA 300
XX DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
XX
XX QY 301 GACGGTGGCGTGTACTGTATACCCAGGATTCACGATTTGAGGATGCTGTTTCATCTAC 360
XX DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
XX
XX QY 361 AAGTCAGTTTCATTGGCGTGAACCTTTCCTCCATGACCTGTATGCAAAAGACACA 420
XX DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
XX
XX QY 421 ATGGCGTGGGAAGCCAGCACTGAGCGTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
XX DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
XX
XX QY 481 ATTCATAGGCTCTGAAGCTGAAAGACGGTGGTTCATTACCTAGTTGAATTCAAAAGTATT 540
XX
Db 161 IleHisLysGAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCAAGAGGCTGTGTGACAGCTTACAGGGTACTACTATCTATCTGACCTCAAACTGGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATAACAAGCCACCAACGAGACTATACATCGTTGAGCAGTATGAAGAACCAGGAGGACGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCATCTGTTCCTT 675
Db 221 HisHisLeuPheLeu 225
XX
XX RESULT 6
XX AAE17540
XX ID AAE17540 standard; protein; 225 AA.
XX AC AAE17540;
XX XX
XX DT 22-APR-2002 (first entry)
XX XX
XX DE Discosoma sp. humanised wild-type Anthozoa protein drFP583.
XX XX
XX KW Fluorescent timer protein; protein movement; translocation; trafficking;
XX KW promoter activity; gene expression; transgenic plant; gene modification;
XX KW protein age; anthozoa protein; drFP583.
XX XX
XX OS Discosoma sp.
XX XX
XX PN WO200196373-A2.
XX XX
XX PD 20-DEC-2001.
XX XX
XX PF 13-JUN-2001; 2001WO-US019097.
XX XX
XX PR 14-JUN-2000; 2000US-0211607P.
XX XX
XX PA (CLON-) CLONTECH LAB INC.
XX XX
XX PI Fradkov AF, Teresikh A;
XX XX
XX DR WPI; 2002-154595/20.
XX XX
XX DR N-PSDB; AAD28207.
XX XX
XX PT New fluorescent timer proteins comprising an emission spectrum that
XX PT changes over time from a first wavelength to a second wavelength, useful
XX PT for monitoring intracellular protein movement, translocation, trafficking
XX PT or stability.
XX XX
XX PS Example 1; Fig 1; 89pp; English.
XX XX
XX CC The invention relates to a fluorescent timer protein having an emission
XX CC spectrum that changes over time after synthesis from a first wavelength
XX CC to a second wavelength. The fluorescent timer proteins are useful in
XX CC monitoring the activity of a promoter, determining the age of a protein,
XX CC identifying an agent that modulates the activity of a promoter and in
XX CC enriching a population of cells comprising a fluorescent timer protein.
XX CC The fluorescent timer proteins are also useful for assessing gene
XX CC expression during development of a multicellular organism or during
XX CC cellular differentiation, in response to a drug or other inducer of
XX CC promoter activity, as a reporter to serve as a read-out of promoter
XX CC activity, monitoring intracellular protein movement or translocation,
XX CC protein trafficking, or protein stability, to investigate temporal
XX CC aspects of the activity of a regulatory element, for determining cell
XX CC fate during development and organ remodelling, in spatial and temporal
XX CC visualisation of newly synthesised proteins and accumulated proteins, and
XX CC in distinguishing between newly formed and pre-existing structures, e.g.
XX CC membrane junctions and extracellular matrix components. The fluorescent
XX CC timer proteins may further be used to investigate where photobleaching
XX CC techniques are employed, as detectable labels, as selectable markers, as
XX CC biosensors in prokaryotic and eukaryotic cells, in protease cleavage
```

CC assays, and as second messenger detectors. The nucleic acids can be used
CC to generate transgenic, non-human plants or animals or site-specific gene
CC modifications in cell lines. The present sequence is *Discooma* sp.
CC humanised wild-type Anthozoa protein drPSP83 used for generating
CC fluorescent proteins
XX
SQ Sequence 225 AA;

Alignment Scores:
Pred. No.: 3,01e-135 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 5 Gaps: 0

US-10-006-922A-11 (1-678) x AAEL17540 (1-225)

Qy 1 ATGAGGTCCTCCAAAGATGTTATCAAGAGTTCATGAGGTTTAAGGTTTCGCATCGAAGGA 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 MetArgSerSerLysAenValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGTCAATGGCCAGCGATTTGAAATAGAGCGAAGAGGAGGGAGGCCATACGAAGGC 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGGTACCAAGGGGGACCTTGGCCATTTCGTTGGGATATT 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTCGCCGACATACCA 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAAGCTGTCATTTCTGAGAGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGCTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
101 AspGlyGlyValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTCAAGTTCATTGGCGTGAACCTTCTTCGATGGACCTGTTATGCAAAAGAGACA 420
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysThr 140
Qy 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTGTGTATGGCGTGTGAAAGGAGAG 480
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAGGCTCTGAAGCTCAAGACGGTGGTCAATTCAGTTGAATTCAAAAGTATT 540
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAAGCCCTGTGCAGCTACCGGTACTACTATGTTGATCTCCAAACTGGAT 600
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATAACAGCCACACGAAGACTATACATCTTTCAGCGAGTATGAAGAACCGAGGAGGC 660
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCCCTT 675
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
221 HisHisLeuPheLeu 225

RESULT 7

AAO18270
ID AAO18270 standard; protein; 225 AA.

XX

AC AAO18270;

XX 26-SEP-2002 (first entry)

XX *Discooma* red fluorescent protein.
DE Yeast; RAD54; promoter; genotoxicity cassette; cytotoxicity cassette;
XX modified yeast strain; environmental pollution.
KW *Discooma* sp.
OS
PN DE10061872-A1.
XX 20-JUN-2002.
PD
XX 12-DEC-2000; 2000DE-01061872.
PF
XX 12-DEC-2000; 2000DE-01061872.
PR
XX (LICH/) LICHTENBERG-FRATE H.
PA
XX Lichtenberg-Frate H;
PI
XX WPI; 2002-539633/58.
DR N-PSDB; AAL47952.

XX Modified yeast strain, useful for detecting toxic compounds in
PT environment, contains integrated cassettes responsive to genotoxic and
PT cytotoxic compounds.
PS Disclosure; Page 21-22; 34pp; German.
XX The present invention relates to a modified yeast strain that contains,
CC integrated stably and functionally in its genome, a genotoxicity cassette
CC and a cytotoxicity cassette, each comprising a promoter and reporter
CC gene, both of which are different in the two cassettes. The modified
CC yeast strain is used to detect environmental pollution, especially
CC genotoxic and/or cytotoxic substances in complex environmental
CC contaminants, especially organic compounds, but also (non-)ionising
CC radiation and chemical carcinogens. Particular applications are in
CC monitoring (waste) water (e.g. as an early warning system), medical
CC toxicology screening and for industrial process control. The present
CC sequence is a marker protein suitable for use in the cassettes of the
CC present invention

SQ Sequence 225 AA;

Alignment Scores:

Pred. No.: 3,01e-135 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 5 Gaps: 0

US-10-006-922A-11 (1-678) x AAO18270 (1-225)

Qy 1 ATGAGGTCCTCCAAAGATGTTATCAAGAGTTCATGAGGTTTAAGGTTTCGCATCGAAGGA 60
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 MetArgSerSerLysAenValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGTCAATGGCCAGCGATTTGAAATAGAGCGAAGGAGGGAGGCCATACGAAGGC 120
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGACCTTTCGATTTGTTGGGATATT 180
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTCGCCGACATACCA 240
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAAGCTGTCATTTCTGAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

QY 301 GACGGTGGCGTCTTACTGTAACCCAGGATTCAGTTTCAGGATGCTGTTTCATCTAC 360
 Db |||||
 QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Db |||||
 QY 361 AAGGTCAGGTTCAATGGCGTGAATCTTCCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
 Db |||||
 QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Db |||||
 QY 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTGATGCGTGTGAAAGGAGAG 480
 Db |||||
 QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Db |||||
 QY 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCACTTACCTAGTTGTAATCAAAAGTATT 540
 Db |||||
 QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 Db |||||
 QY 541 TACATGGCAAGAACGCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT 600
 Db |||||
 QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 Db |||||
 QY 601 ATAAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAAGCGGAGCGC 660
 Db |||||
 QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Db |||||
 QY 661 CACCATCTGTTCTCTT 675
 Db |||||
 QY 221 HisHisLeuPheLeu 225

RESULT 8

ID AAE34962 standard; protein; 225 AA.
 XX AC AAE34962;

XX 28-MAY-2003 (first entry)

XX Discosoma species red fluorescent protein (RFP).

XX Phosphorylation indicator; fluorescent protein; detection; phosphatase;
 KW kinase; red fluorescent protein; RFP.

XX Discosoma sp.

XX W0200295058-A2.

XX 28-NOV-2002.

XX 24-MAY-2002; 2002WO-US016955.

XX 24-MAY-2001; 2001US-00865291.

XX (REGC) UNIV CALIFORNIA.

XX Tsien RY, Ting AY, Zhang J;

XX WPI; 2003-149474/14.

XX N-PSDB; AAD53432.

XX Novel chimeric phosphorylation indicators, useful for detecting
 PT kinase/phosphatase in samples, has donor molecule, phosphorylatable
 PT domain, phosphoaminoacid binding domain, and acceptor molecule, in
 PT operative linkage.

XX Disclosure; Col 65-66; 38pp; English.

XX The present invention relates to chimeric phosphorylation indicators
 CC comprising a phosphorylation polypeptide and a fluorescent protein or
 CC operative linkage, a donor molecule, a phosphorylatable domain, a
 CC phosphoaminoacid binding domain (PAAAD) and an acceptor molecule. The
 CC phosphorylation indicators of the invention are useful for detecting
 CC kinases or phosphatases in a biological sample. They are also useful in
 CC high throughput analysis e.g. for detecting a kinase inhibitor or

CC phosphatase inhibitor. The present sequence is Discosoma species red
 CC fluorescent protein (RFP) used in the invention
 XX
 SQ Sequence 225 AA;

Alignment Scores:
 Pred. No.: 3,01e-135 Length: 225
 Score: 1214.00 Matches: 225
 Percent Similarity: 100.00% Conservativity: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.92% Indels: 0
 DB: Gaps: 0

US-10-006-922A-11 (1-678) x AAE34962 (1-225)

QY 1 ATGAGGTCTTCCAGGAATGTTATCAAGGAGTTCATGAGGTTTAAGGTTCCATGGAAGGA 60
 Db |||||
 QY 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
 Db |||||
 QY 61 ACGTCAATGGCAGCAGGTTTGAATAGAGGGCAGGAGGAGGGCCATACGAAGGC 120
 Db |||||
 QY 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
 Db |||||
 QY 121 CACAATACCCGTAAAGCTTAAGGTAAACCAAGGGGACCTTTGCCATTTGCTGGGATATT 180
 Db |||||
 QY 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
 Db |||||
 QY 181 TTGTCACACAAATTTCAAGTATGGAAGCAAGGTATATGTCGAAGCACCTTGGCGACATACCA 240
 Db |||||
 QY 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
 Db |||||
 QY 241 GACTATAAAGCTGTCATTTCTCTGAAGGATTTAAATGGGAAAGGGTCATGAATTTGAA 300
 Db |||||
 QY 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
 Db |||||
 QY 301 GACGGTGGCGTCTTACTGTAACCCAGGATTTCCAGATTTCAGGATGCTGTTTCATCTAC 360
 Db |||||
 QY 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
 Db |||||
 QY 361 AAGTCAAGTTCATTTGGCGTGAATCTTCCCTCCGATGGACCTGTTATGCAAAAGAGACA 420
 Db |||||
 QY 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
 Db |||||
 QY 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTGATGCGTGTGAAAGGAGAG 480
 Db |||||
 QY 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
 Db |||||
 QY 481 ATTCATAGGCTCTGAAGCTGAAGAGCGTGGTCACTTACCTAGTTGTAATCAAAAGTATT 540
 Db |||||
 QY 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
 Db |||||
 QY 541 TACATGGCAAGAACGCTGTGCAGCTACCGGGTACTACTATGTTGACTCCAAACTGGAT 600
 Db |||||
 QY 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
 Db |||||
 QY 601 ATAAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAAGCGGAGCGC 660
 Db |||||
 QY 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
 Db |||||
 QY 661 CACCATCTGTTCTCTT 675
 Db |||||
 QY 221 HisHisLeuPheLeu 225

RESULT 9

ID ADC24126 standard; protein; 225 AA.
 XX AC ADC24126;

XX 18-DEC-2003 (first entry)

XX Discosoma wild-type red fluorescent protein.


```
XX TSien RY, Zacharias DA, Baird GS;
XX WPI; 2003-802418/75.
XX DR N-PSDB; AAD61969.
XX
XX Fluorescent proteins containing a mutation that reduces or eliminates its
XX ability to oligomerize which gives more reliable fluorescence resonance
XX energy transfer results and are useful to detect molecule interaction,
XX enzymes, or sample pH.
XX
XX PS Claim 10; Page 30-31; Opp; English.
XX
XX CC The invention relates to a non-oligomerising fluorescent protein
XX containing a mutation that reduces or eliminates its ability to
XX oligomerise. The fluorescent protein gives more reliable fluorescence
XX resonance energy transfer results and are useful to detect molecule
XX interaction, enzymes, or sample pH. These are also used to identify
XX agents or conditions that regulate expression of control sequences. The
XX present sequence is Discosoma sp. red fluorescent protein (RFP)
XX
XX SQ Sequence 225 AA;
XX
XX Alignment Scores:
XX Pred. No.: 3,01e-135 Length: 225
XX Score: 1214.00 Matches: 225
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 99.92% Indels: 0
XX DB: 7 Gaps: 0
XX
XX US-10-006-922A-11 (1-678) x ABW09018 (1-225)
XX
XX QY 1 ATGAGGTCTTCCAAGATTTATCAAGAGTTCATGAGTTTAAGGTTTCATGGAAGCA 60
XX DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
XX
XX QY 61 ACGTCAATGGCCGAGTTTGAATAGAGGGGAGGAGGGGAGGCCATACGAAGGC 120
XX DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
XX
XX QY 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGACCTTTGCCATTGCTTGGGATATT 180
XX DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
XX
XX QY 181 TTGTCAACCAATTTCAAGTATGGAAGCAAGGTATATGTCACGACCTCGCCGACATACCA 240
XX DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
XX
XX QY 241 GACTATAAAGCTGTGCTATTCTCGAAGGATTTAAATGGGAAGGGTCATGAATTTGAA 300
XX DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
XX
XX QY 301 GAGGTGGCGTCTACTGTAAACCCAGGATTCAGATTTCAGAGTGGCTGCTTTCATCTAC 360
XX DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
XX
XX QY 361 AAGGTCAAGTTCATFTGCGGTGAATTTCCCTTCCGATGGACCTGTATGCAAAAGAGACA 420
XX DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
XX
XX QY 421 ATGGGCTGGGAAGCCAGCTAGCGTGTGTTATCTCTCGTATGCGGTGTTGAAAGGAGAG 480
XX DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
XX
XX QY 481 ATTTCATAGGCTCTGAAGCTGAAGACGGTGGTCATTACCTAGTTCGATTCGAAAGTATT 540
XX DB 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
XX
XX QY 541 TACATGGCAAGAGCCCTGTGACGCTACCGAGGTACTACTATGTTGACTCCAAACTGGAT 600
XX DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
XX
XX QY 601 ATAAAGCCACCAACGAGCTATCAATCGTTGACGAGTATGAAGAACCGAGGACGC 660
XX DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
XX
XX QY 661 CACCATCTGTTCCTT 675
XX DB 221 HisHisLeuPheLeu 225
XX
XX RESULT 11
XX ADF70403
XX ID ADF70403 standard; protein; 225 AA.
XX
XX AC ADF70403;
XX
XX DT 12-FEB-2004 (first entry)
XX
XX DE Discosoma wild-type GFP variant protein SeqID26.
XX
XX KW ligand; orphan receptor protein; fusion protein; fluorescent protein;
XX cell expression; green fluorescent protein; GFP; GFP-1; wild-type GFP;
XX GFPuv; Enhanced GFP; EGFP.
XX
XX OS Discosoma sp.
XX
XX PN WO2003071272-A1.
XX
XX PD 28-AUG-2003.
XX
XX PF 21-FEB-2003; 2003WO-JP001901.
XX
XX PR 22-FEB-2002; 2002JP-00045728.
XX
XX PR 23-JUL-2002; 2002JP-00213949.
XX
XX PR 11-OCT-2002; 2002JP-00298237.
XX
XX (TAKE ) TAKEDA CHEM IND LTD.
XX
XX PI Hinuma S, Fujii R, Ogi K, Komatsu H, Kawamata Y, Hosoya M;
XX
XX DR WPI; 2003-697654/66.
XX
XX DR N-PSDB; ADF70404.
XX
XX PT Transformation of cells with a fusion protein of an orphan receptor
XX protein with a fluorescent protein useful for identification of ligands
XX to the orphan receptor.
XX
XX PS Disclosure; SEQ ID NO 26; 594pp; Japanese.
XX
XX CC This invention relates to a novel method of identifying ligands to an
XX orphan receptor protein which comprises transforming cells with DNA
XX encoding a fusion protein of the orphan receptor with a fluorescent
XX protein, so that the fusion protein is expressed in the cells (or cell
XX membranes isolated from them) and contacting the cells with the potential
XX ligand to be tested. A suitable fluorescent protein for incorporation in
XX the fusion protein is green fluorescent protein (GFP), for example GFP-1,
XX wild-type GFP, GFPuv or Enhanced GFP (EGFP). The method is useful for the
XX identification of ligands binding to an orphan receptor protein.
XX
XX SQ Sequence 225 AA;
XX
XX Alignment Scores:
XX Pred. No.: 3,01e-135 Length: 225
XX Score: 1214.00 Matches: 225
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 99.92% Indels: 0
XX DB: 7 Gaps: 0
XX
XX US-10-006-922A-11 (1-678) x ADF70403 (1-225)
XX
XX QY 1 ATGAGGTCTTCCAAGATTTATCAAGAGTTCATGAGTTTAAGGTTTCATGGAAGCA 60
XX DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
```



```
Query Match: 99.92% Indels: 0
DB: 7 Gaps: 0
US-10-006-922A-11 (1-678) x ADH34489 (1-225)
QY 1 ATGAGGCTCTCCAGAATGTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGGA 60
DB 1 MetArgSerSerLysAsnValIleGlyGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACGGTCATGGGACGAGTTTGAATAGAGGCGAAGGAGGAGGAGGCCATACGAAGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGlyGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGAGCTTTGCCATTTGCGGATATT 180
DB 41 HisAsnThrValLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 60
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTTCGCGCATACCA 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTATATAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTGAA 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGTGGCGCTGTTACTGTAAACCCAGGATTTCCAGATTTCCAGGATGCTGTTTCATCTAC 360
DB 101 AspGlyGlyValValThrValThrValThrValThrValThrValThrValThrValThr 120
QY 361 AAGGTCAAGTTCATTCGCGTGAACCTTTCTTCGATGGAGCTTTCTTCGATGGAGCTTT 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCGTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGCTCATCTAGTTGAATTCAAAGTATT 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCAAGAGCCTGTCAGCTACCGGCTACTATGTTGACTCCAACTCGAT 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATAAACAGCCACACGAGAGCTATACAACTCTTCAGCAGTATGAAGAACCAGGAGCAGC 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCATCTGTTCTTT 675
DB 221 HisHisLeuPheLeu 225
RESULT 13
ID ADL46203 standard; protein; 225 AA.
XX AC
XX ADL46203;
XX DT 20-MAY-2004 (first entry)
XX DE Discosoma red fluorescent (DeRed) protein.
XX KW red fluorescent protein; DeRed; fluorescence; red wavelength;
XX KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX OS Discosoma sp.
XX FN WO2003086446-A1.
XX PD 23-OCT-2003.
XX PF 09-APR-2003; 2003WO-US010879.
```

XX 10-APR-2002; 2002US-00121258.
PR 29-JUL-2002; 2002US-00209208.
XX (REGC) UNIV CALIFORNIA.
XX Tsien RY, Campbell RE, Baird GS;
XX WPI; 2003-845265/78.
DR N-PSDB; ADL46204.
XX New monomeric and dimeric Anthozoan fluorescent protein variants with
PT reduced propensity to oligomerize, and encoding polynucleotides, useful
PT in molecular biology, e.g. in immunoassays or in tracking protein
PT movement in cells.
XX Claim 1; SEQ ID NO 1; 166pp; English.
XX The invention relates to a polynucleotide sequence encoding a Discosoma
CC red fluorescent protein (DeRed) variant having a reduced propensity to
CC oligomerize. The protein variant comprises one or more amino acid
CC substitutions at the AB and/or AC interface(s) of the wild-type DeRed
CC sequence, where the substitutions result in reduced propensity of the
CC DeRed variant to form tetramers and where the variant displays detectable
CC fluorescence of at least one red wavelength. The composition and methods
CC are useful in producing red fluorescent proteins having reduced
CC propensity for oligomerization, especially tetramerization. The protein
CC may be used in molecular biology and in other scientific applications,
CC such as in immunoassays or hybridization assays, or in tracking the
CC movement of proteins in cells. This sequence corresponds to the DeRed
CC protein.
XX SQ Sequence 225 AA;
Alignment Scores:
Pred. No.: 3,01e-135 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 7 Gaps: 0
US-10-006-922A-11 (1-678) x ADL46203 (1-225)
QY 1 ATGAGGCTCTCCAGAATGTTATCAAGGAGTTTCATGAGTTTAAAGGTTTCGCATGGAAGGA 60
DB 1 MetArgSerSerLysAsnValIleGlyGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACGGTCATGGGACGAGTTTGAATAGAGGCGAAGGAGGAGGCCATACGAAGGC 120
DB 21 ThrValAsnGlyHisGluPheGluIleGlyGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGAGCTTTGCCATTTGCGGATATT 180
DB 41 HisAsnThrValLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 60
QY 181 TTGTCAACCAATTTTCAGTATGGAAGCAAGGTATATGTCGAAGCACCTTCGCGCATACCA 240
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTATATAAAGCTGTCATTTCTGAAGGATTTAAATGGGAAGGGTCATGAACCTTGAA 300
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGGTGGCGCTGTTACTGTAAACCCAGGATTTCCAGATTTCCAGGATGCTGTTTCATCTAC 360
DB 101 AspGlyGlyValValThrValThrValThrValThrValThrValThrValThrValThr 120
QY 361 AAGGTCAAGTTCATTCGCGTGAACCTTTCTTCGATGGAGCTTTCTTCGATGGAGCTTT 420
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCGTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGCGGTGTTGAAAGGAGAG 480
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATTCAATAGGCTCTGAAGCTGAAGAGCGGTGCTCATCTAGTTGAATTCAAAGTATT 540
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCAAGAGCCTGTCAGCTACCGGCTACTATGTTGACTCCAACTCGAT 600
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
QY 601 ATAAACAGCCACACGAGAGCTATACAACTCTTCAGCAGTATGAAGAACCAGGAGCAGC 660
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCATCTGTTCTTT 675
DB 221 HisHisLeuPheLeu 225
RESULT 13
ID ADL46203 standard; protein; 225 AA.
XX AC
XX ADL46203;
XX DT 20-MAY-2004 (first entry)
XX DE Discosoma red fluorescent (DeRed) protein.
XX KW red fluorescent protein; DeRed; fluorescence; red wavelength;
XX KW oligomerization; tetramerization; immunoassay; hybridization assay.
XX OS Discosoma sp.
XX FN WO2003086446-A1.
XX PD 23-OCT-2003.
XX PF 09-APR-2003; 2003WO-US010879.

Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCAATAGGCTCTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAATCAAAAGATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGCCAAAGAGCCTGTGCACCTACCGGGTACTATGTTGACTCCAAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACAAAGCCACACGAGACTATACAAATCGTCGAGCAGTATGAAGAACCAGGGACGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 14
ADN33979
ID ADN33979 standard; protein; 225 AA.
XX AC ADN33979;
XX DT 01-JUL-2004 (first entry)
XX DE Wild-type DsRED protein.
XX KW Chidarian; fluorescence resonance energy transfer; FRET; wild-type DsRED.
XX OS Discosoma sp.
XX PN W02003054158-A2.
XX PD 03-JUL-2003.
XX PF 18-DEC-2002; 2002WO-US040539.
XX PR 19-DEC-2001; 2001US-0341723P.
XX PA (UYCH-) UNIV CHICAGO.
XX PI Bevis B, Glick B;
XX PT WPI; 2003-569236/53.
XX DR N-PSDB; ADN33978.
XX PT Novel nucleic acid encoding a rapidly maturing chromo- or fluorescent mutant of a Chidarian chromo- or fluorescent protein or its mutant, useful for applications involving chromo- or fluorescent proteins.
XX PS Claim 8; SEQ ID NO 2; 65pp; English.
XX The present invention relates to nucleic acid that encodes a rapidly maturing chromo or fluorescent mutant of a Chidarian chromo- or fluorescent protein or its mutant. The protein is useful in applications involving nucleic acid encoding a chromo- or fluorescent protein and is useful for producing a chromo and/or fluorescent protein which involves growing the cell, whereby the protein is expressed, and isolating the protein substantially free of other proteins. The protein is useful in applications involving chromo- or fluorescent protein and is useful as PCR primers, hybridization probes, etc. The expression cassettes are useful for synthesizing related proteins. The chromoproteins are useful as coloring agents which are capable of imparting color or pigment to a particular composition of matter e.g. food compositions, pharmaceuticals, cosmetics, living organisms, e.g., animals and plants. The chromoproteins may also find use as labels in analyte detection assays, e.g. assays for biological analytes of interest and as selectable markers in recombinant DNA applications, e.g. the production of transgenic cells and organisms. The fluorescent proteins find use in a variety of different applications, e.g. in fluorescence resonance energy transfer (FRET) applications, as biosensors in prokaryotic and eukaryotic cells, in applications involving

CC the automated screening of arrays of cells expressing fluorescent reporting groups by using microscopic imaging and electronic analysis, as second messenger detectors, and in fluorescence activated cell sorting applications and as in vivo marker in animals. The fluorescent proteins CC also find use in protease cleavage assays. The proteins can also be used CC is assays to determine the phospholipid composition in biological CC membranes and as a fluorescent timer. The present sequence represents the CC wild-type DsRED.
XX Sequence 225 AA;
SQ
Alignment Scores: 3.01e-135 Length: 225
Pred. No.: 1214.00 Matches: 225
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 99.92% Indels: 0
Query Match: 7 Gaps: 0
DB:
US-10-006-922A-11 (1-678) x ADN33979 (1-225)
Qy 1 ATGAGGTCTTCCAAGAATGTTATCAAGGAGTTTCATGAGTTTAAAGTTCCGATCGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGGCACGAGTTTGAATAGAACGAGGAGGAGGAGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAGGTAACCAAGGGGGACCTTTGCCATTTGCTCGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTatPheIle 60
Qy 181 TTGTACCACAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAGCTGTCTTCCTGAAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGTCTTACTGTAAACCCAGGATTCAGTTTCAGATGGCTGTTTCACTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGTCAAGTTTCATTCGCGTGAACCTTTCCTCCGATGGACCTCTTATGCAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCCTCGTATGGCGTGTGAAAGAGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCAATAGGCTCTGAAGCTGAAGACGGTGGTCAATACCTAGTTGAATCAAAAGATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGCCAAAGAGCCTGTGCACCTACCGGGTACTATGTTGACTCCAAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACAAAGCCACACGAGACTATACAAATCGTTGAGCAGTATGAAGAACCAGGGACGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225
RESULT 15
AD136421
ID AD136421 standard; protein; 225 AA.
XX

THIS PAGE BLANK (US-10)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 30, 2005, 15:54:05 ; Search time 137.5 Seconds
(without alignments)
3804.643 Million cell updates/sec

Title: US-10-006-922A-11

Perfect score: 1215
Sequence: 1 atgaggcttccaagaatgt.....gccaccatctgtccctttaa 678

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1722976 seqs, 385795295 residues

Total number of hits satisfying chosen parameters: 3445952

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2_1/usPTO spool_p/US10006922/runat_30062005_110554_9471/app_query.fasta_1.839
-DB=Published Applications AA -QFMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPECL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10006922 @CGN 1 1 199 @runat 30062005_110554_9471
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FEAPOP=6 -FEAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US10E_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US11A_PUBCOMB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
21: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
22: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1214	99.9	225	9	US-09-999-745-67 Sequence 67, Appl
2	1214	99.9	225	10	US-09-866-538-12 Sequence 12, Appl
3	1214	99.9	225	10	US-09-794-308-12 Sequence 12, Appl
4	1214	99.9	225	10	US-09-865-291-12 Sequence 12, Appl
5	1214	99.9	225	13	US-10-006-922-12 Sequence 12, Appl
6	1214	99.9	225	14	US-10-081-864-8 Sequence 8, Appl
7	1214	99.9	225	14	US-10-121-258-1 Sequence 1, Appl
8	1214	99.9	225	14	US-10-315-920-2 Sequence 2, Appl
9	1214	99.9	225	15	US-10-132-067-4 Sequence 4, Appl
10	1214	99.9	225	15	US-10-370-570-56 Sequence 56, Appl
11	1214	99.9	225	15	US-10-406-618-32 Sequence 32, Appl
12	1214	99.9	225	16	US-10-433-640-13 Sequence 13, Appl
13	1214	99.9	225	16	US-10-311-030-7 Sequence 7, Appl
14	1214	99.9	225	16	US-10-845-484-3 Sequence 3, Appl
15	1214	99.9	225	16	US-10-885-988-12 Sequence 12, Appl
16	1214	99.9	225	17	US-10-656-029-22 Sequence 22, Appl
17	1214	99.9	225	17	US-10-857-622-12 Sequence 12, Appl
18	1214	99.9	225	17	US-10-505-486-26 Sequence 26, Appl
19	1214	99.9	487	15	US-10-343-977-1 Sequence 1, Appl
20	1214	99.9	487	17	US-10-343-977-1 Sequence 2, Appl
21	1214	99.9	506	15	US-10-343-977-2 Sequence 2, Appl
22	1214	99.9	506	17	US-10-343-977-3 Sequence 3, Appl
23	1214	99.9	547	15	US-10-343-977-3 Sequence 3, Appl
24	1214	99.9	547	17	US-10-343-977-3 Sequence 9, Appl
25	1210	99.6	226	16	US-10-311-030-9 Sequence 2, Appl
26	1210	99.6	240	14	US-10-152-296-2 Sequence 2, Appl
27	1210	99.6	240	16	US-10-739-656-2 Sequence 2, Appl
28	1210	99.6	545	14	US-10-214-932-52 Sequence 52, Appl
29	1210	99.6	548	14	US-10-214-932-76 Sequence 76, Appl
30	1207	99.3	225	14	US-10-121-258-20 Sequence 20, Appl
31	1207	99.3	225	14	US-10-315-920-4 Sequence 4, Appl
32	1196	98.4	236	16	US-10-314-936-2 Sequence 2, Appl
33	1196	98.4	236	16	US-10-314-936-4 Sequence 4, Appl
34	1196	98.4	236	19	US-11-021-014-2 Sequence 2, Appl
35	1196	98.4	236	19	US-11-021-014-4 Sequence 4, Appl
36	1195.5	98.4	226	16	US-10-724-178-12 Sequence 12, Appl
37	1191	98.0	225	14	US-10-315-920-6 Sequence 6, Appl
38	1186	97.6	225	15	US-10-442-148A-7 Sequence 7, Appl
39	1186	97.6	239	15	US-10-442-148A-8 Sequence 8, Appl
40	1160	95.5	225	14	US-10-121-258-4 Sequence 4, Appl
41	1153	94.9	225	14	US-10-121-258-24 Sequence 24, Appl
42	1132	93.2	225	15	US-10-423-688A-41 Sequence 41, Appl
43	1121	92.3	226	14	US-10-121-258-6 Sequence 6, Appl
44	1119	92.1	225	13	US-10-006-922-44 Sequence 44, Appl
45	1119	92.1	225	14	US-10-081-864-12 Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-09-999-745-67
; Sequence 67, Application US/09999745
; Patent NO. US20020157120A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Baird, Geoffrey
; TITLE OF INVENTION: CIRCULARLY PERMUTED FLUORESCENT PROTEIN INDICATORS
; FILE REFERENCE: REGEN1470-1
; CURRENT APPLICATION NUMBER: US/09/999,745
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: 09/316,920
; PRIOR FILING DATE: 1999-05-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 67
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.

US-09-999-745-67

Alignment Scores:

Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 9 Gaps: 0

US-10-006-922A-11 (1-678) x US-09-999-745-67 (1-225)

```
QY 1 ATGAGGCTCTCCAAAGAAATGTTATCAAGAGTTTCATGAGTTTAAGGTTTCGCATGCAAGGA 60
   |||
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

QY 61 ACGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGAGGGAGGCCATACGAAGGC 120
   |||
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40

QY 121 CACATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTCCTGGGATATT 180
   |||
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

QY 181 TTGTCAACACAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240
   |||
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

QY 241 GACTATAAAAGCTGTCTATTTCTGAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 300
   |||
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluA-gValMetAsnPheGlu 100

QY 301 GACGGTGGCGTCTACTGTAAACCCAGGATTCAGTTTGACAGGATGCTGTTTCATCTAC 360
   |||
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

QY 361 AAGTCAAGTTCATTTGGCGTGAACCTTTCTTCCATGGACCTGTATGCAAAAGAGACA 420
   |||
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

QY 421 ATGGCTGGGAAGCCAGCAGCTAGCGTTGTATCTCGTGTAGTGGCGTGTGAAGAGAGAG 480
   |||
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

QY 481 ATTCATAGGCTCTGAGCTCAAGAGCGTGGTCACTTACCTAGTGAATTCATAAGATTT 540
   |||
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

QY 541 TACATGGCAAGAACCCCTGTGCAGCTACCGGGTACTATGTTGACTCCAAACTGGAT 600
   |||
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

QY 601 ATAAACAGCCACACGAAGACTATACAACTCTTGAGCAGGTATGAAGAACCGAGGAGCGC 660
   |||
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220

QY 661 CACCATCTGTTCTCTT 675
   |||
DB 221 HisHisLeuPheLeu 225
```

RESULT 2

US-09-866-538-12
; Sequence 12, Application US/09866538
; Publication No. US20030032088A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/09/866,538
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 12

; LENGTH: 225

; TYPE: PRT

; ORGANISM: Discosoma sp.

US-09-866-538-12

Alignment Scores:

Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 10 Gaps: 0

US-10-006-922A-11 (1-678) x US-09-866-538-12 (1-225)

```
QY 1 ATGAGGCTCTCCAAAGAAATGTTATCAAGGAGTTTCATGAGGTTTAAGGTTTCGCATGGAAGGA 60
   |||
DB 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

QY 61 ACGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGAGGGAGGCCATACGAAGGC 120
   |||
DB 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40

QY 121 CACATACCGTAAAGCTTAAGGTAACCAAGGGGGGACCTTTGCCATTTCCTGGGATATT 180
   |||
DB 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

QY 181 TTGTCAACACAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240
   |||
DB 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

QY 241 GACTATAAAAGCTGTCTATTTCTGAAGGATTTAAATGGGAAAGGTCATGAACCTTTGAA 300
   |||
DB 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

QY 301 GACGGTGGCGTCTACTGTAAACCCAGGATTCAGTTTGACAGGATGCTGTTTCATCTAC 360
   |||
DB 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

QY 361 AAGTCAAGTTCATTTGGCGTGAACCTTTCTTCCATGGACCTGTATGCAAAAGAGACA 420
   |||
DB 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

QY 421 ATGGCTGGGAAGCCAGCAGCTAGCGTTGTATCTCGTGTAGTGGCGTGTGAAGAGAGAG 480
   |||
DB 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160

QY 481 ATTCATAGGCTCTGAGCTCAAGAGCGTGGTCACTTACCTAGTGAATTCATAAGATTT 540
   |||
DB 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

QY 541 TACATGGCAAGAACCCCTGTGCAGCTACCGGGTACTATGTTGACTCCAAACTGGAT 600
   |||
DB 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

QY 601 ATAAACAGCCACACGAAGACTATACAACTCTTGAGCAGGTATGAAGAACCGAGGAGCGC 660
   |||
DB 201 IleThrSerHisAsnGluAspTyrThrIleValGlnTyrGluArgThrGluGlyArg 220

QY 661 CACCATCTGTTCTCTT 675
   |||
DB 221 HisHisLeuPheLeu 225
```

RESULT 3

US-09-794-308-12
; Sequence 12, Application US/09794308
; Publication No. US20030170911A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: ZACHARIAS, David

; APPLICANT: BAIRD, Geoffrey
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGN1530
; CURRENT APPLICATION NUMBER: US/09/794,308
; CURRENT FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-794-308-12

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 10 Gaps: 0

US-10-006-922A-11 (1-678) x US-09-794-308-12 (1-225)

Qy	1	ATGAGGCTTCCAGAGTTCATCAAGAGTTCATGAGGTTTAAGGTTCCGATCGAAGGA	60
Db	1	MetArgSerLysAsnValIleGluPheMetArgPheLysValArgMetGluGly	20
Qy	61	ACGGTCAATGGCCAGAGTTTGAATAGAACGAGGAGGAGGCCATACGAAGGC	120
Db	21	ThrValasnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Qy	121	CACAAATCCGTAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTGCTGGGATATT	180
Db	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Qy	181	TTGTCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTCCGACATACCA	240
Db	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Qy	241	GACTATAAAAGCTGTCAATTTCTGAAGGATTTAAATGGGAAAGGGTCATCAACTTTGAA	300
Db	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Qy	301	GACGCTGGCGTCTACTGTAAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC	360
Db	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Qy	361	AAGTCAAGTTCAATGGCGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAGACA	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Qy	421	ATGGCTGGGAGCCAGCAGCTGAGCGTTTGTATCTCTGATGGCGTTCGAAAGGAGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Qy	481	ATTCAATAGGCTCTGAAGCTGAAGACCGTGGTTCATTAACCTAGTTGAATTCAAAGATT	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
Qy	541	TACATGGCAAGAACCTGTGAGCTACACGGGTACTACTATGTTGACTCCAAACTGGAT	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Qy	601	ATACAGCCACACGAGACTATCAATCTGTGAGCAGTATGAAGAACCGAGGGAGCG	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Qy	661	CACCATCTGTTCTCTT	675
Db	221	HisHisLeuPheLeu	225

RESULT 4

US-09-865-291-12

; Sequence 12, Application US/09865291
; Publication No. US20030186229A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TSJEN, Roger
; APPLICANT: TING, Alice
; APPLICANT: ZHANG, Jin
; TITLE OF INVENTION: EMISSION RATIONETRIC INDICATORS OF PHOSPHORYLATION
; FILE REFERENCE: REGN1550
; CURRENT APPLICATION NUMBER: US/09/865,291
; CURRENT FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-09-865-291-12

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 10 Gaps: 0

US-10-006-922A-11 (1-678) x US-09-865-291-12 (1-225)

Qy	1	ATGAGGCTTCCAGAGTTCATCAAGAGTTCATGAGGTTTAAGGTTCCGATCGAAGGA	60
Db	1	MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Qy	61	ACGGTCAATGGCCAGAGTTTGAATAGAACGAGGAGGAGGCCATACGAAGGC	120
Db	21	ThrValasnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Qy	121	CACAAATCCGTAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTTGCTGGGATATT	180
Db	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Qy	181	TTGTCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTCCGACATACCA	240
Db	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Qy	241	GACTATAAAAGCTGTCAATTTCTGAAGGATTTAAATGGGAAAGGGTCATCAACTTTGAA	300
Db	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Qy	301	GACGCTGGCGTCTACTGTAAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC	360
Db	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Qy	361	AAGTCAAGTTCAATGGCGTGAACCTTCTCCGATGGACCTGTTATGCAAAAGAGACA	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Qy	421	ATGGCTGGGAGCCAGCAGCTGAGCGTTTGTATCTCTGATGGCGTTCGAAAGGAGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Qy	481	ATTCAATAGGCTCTGAAGCTGAAGACCGTGGTTCATTAACCTAGTTGAATTCAAAGATT	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle	180
Qy	541	TACATGGCAAGAACCTGTGAGCTACACGGGTACTACTATGTTGACTCCAAACTGGAT	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Qy	601	ATACAGCCACACGAGACTATCAATCTGTGAGCAGTATGAAGAACCGAGGGAGCG	660
Db	201	IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220

```
Qy 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 5
US-10-006-922-12
; Sequence 12, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey A
; APPLICANT: Fradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tersikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma species
US-10-006-922-12

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 13 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-006-922-12 (1-225)
Qy 1 ATGAGGCTCTTCCAAGATGTTTATCAAGGAGTTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGGAAGAGAGGGAGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluLeuGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACATACCGTAAAGCTTAAGCTAACCAAGGGGGACCTTGGCATTGCTTGGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAGACCCCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATATAAAGCTGTCTATTTCTCAAGGATTTAAATGGGAAGGGGTTCATGAATTTCAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCTCCTTACTGTACCCAGGATTCAGTTTGCAGGATGGCTGTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

Qy 361 AAGGTCAAAGTTTCATTCGGCTGAACCTTTCTCTCCGATGGACCTGTTATGCAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGGTGGGAAGCCAGCACTGAGCGTTTGTATCTCTCGTATGCGGTGTTGAAAGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCAATAGGCTCTGAAGCTGAAGACGGTGGTTCATTACTAGTTCAATTCAAAGATATT 540
Db 161 IleHisLysAlaLeuLysLysAspGlyGlyHisIleLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAACGCTGTGACAGCTACACAGGCTACTACTATCTTCACTCCAACTGGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATAACAGCCACCAAGCAACTATACAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 6
US-10-081-864-8
; Sequence 8, Application US/10081864
; Publication No. US20030022287A1
; GENERAL INFORMATION:
; APPLICANT: Lukanov, Sergey
; APPLICANT: Yanyanov, Konstantin
; APPLICANT: Yanyanov, Yuriy
; APPLICANT: Savitsky, Alexandr
; APPLICANT: Fradkov, Arcady
; TITLE OF INVENTION: No. US20030022287A1 Aggregating Fluorescent Proteins and
; TITLE OF INVENTION: Methods for Using the Same
; FILE REFERENCE: CLON-067
; CURRENT APPLICATION NUMBER: US/10/081,864
; CURRENT FILING DATE: 2002-06-19
; PRIOR APPLICATION NUMBER: 10/006,922
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/270,983
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-081-864-8

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 14 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-081-864-8 (1-225)
Qy 1 ATGAGGCTCTTCCAAGATGTTTATCAAGGAGTTCATGAGTTTAAAGTTTCGCATGGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAGCGGAAGAGAGGGAGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluLeuGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACATACCGTAAAGCTTAAGCTAACCAAGGGGGACCTTGGCATTGCTTGGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
```

181	QY	TTGTCTACCA	CAATTTTCAGTATGG	AAAGGTATATGTCAAGCA	CCCTGCCGACATACCA	240
61	Db	LeuSerProGlnPheGln	TyrGlySerIysValTyrVall	YsHisProAlaAspIlePro	80	
241	QY	GACTATAAAAGCTGCT	CAATTTCCCTCAAGGAT	TAAATGGGNAAGGCTCAT	GAACTTTGA	300
81	Db	AspTyrLysLysLeu	SerPheProGluGlyPhe	LysTrpGluArgValMet	AsnPheGlu	100
301	QY	GACGGTGGCGTTCGT	TACTGTAAACCAGGAT	TCCAGTTTGAGGATGGCT	GTTCATCTAC	360
101	Db	AspGlyGlyValVal	ThrValThrCln	AspSerSerLeuGln	AspGlyCysPheIleTyr	120
361	QY	AAGGTCAGTTCAT	TGGCGTGAACCTTTC	TCCGATGACCTGT	TATGC AAAAAGAC	420
121	Db	LysValLysPheIle	GlyValAsnPhePro	SerAspGlyProVal	MetGlnLysLysThr	140
421	QY	ATGGCGTGGGAAGC	CAGCACTGAGCGTT	TGTATCCTCGTGA	TGGCGTGTGAAAGGAG	480
141	Db	MetGlyTrpGluAla	SerThrGluArgLeu	TyrProArgAspGly	ValLeuLysGlyGlu	160
481	QY	ATTCATAGGCTCTA	AGCTGAAGACGGTGGT	CAATTACCTAGT	TGATTCGAATTC	540
161	Db	IleHisLysAlaLeu	LysLeuLysAspGly	GlyHisTyrLeuVal	GluPheLysSerIle	180
541	QY	TACATGGCAAGAAGC	CTGTGCAGCTAC	CGGGTACTATGT	TGACTCCAAACCTGGAT	600
181	Db	TyrMetAlaLysLys	ProValGlnLeuPro	GlyTyrTyrVal	AspSerLysLeuAsp	200
601	QY	ATAACAAGCCACAAC	GAAGACTATACAAT	CTGTGAGCAGTAT	CAAAAGAACCGGAGGC	660
201	Db	IleThrSerHis	AsnGluAspTyrThr	IleValGluGlnTyr	GluArgThrGluGlyArg	220
661	QY	CACCATCTGTTCC	TTT	675		
221	Db	HisHisLeuPhe	Leu	225		

RESULT.7

```

US-10-121-258-1
; Sequence 1, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; TITLE OF INVENTION: PROTEIN VARIANTS AND METHODS FOR MAKING SAME
; FILE REFERENCE: UC083.1CP2CP1
; CURRENT APPLICATION NUMBER: US/10/121,258
; CURRENT FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER:..09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(225)
; OTHER INFORMATION: wild-type DsRed
US-10-121-258-1

```

Alignment Scores:	
Pred. No.:	2,99e-128
Score:	1214.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	99.92%
DB:	14
	Length:
	Matches:
	Mismatches:
	Indels:
	Gaps:

US-10-006-922A-11 (1-678) x US-10-121-258-1 (1-225)

Qy	1	ATGAGGCTTCCCAAGATGTTATCAACGAGGTTCATGAGGTTTAAGGTTTCGCATATGCAAGGA	60
Db	1	MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly	20
Qy	61	ACGGTCAATGGGCACGAGTGTGAAATAGAAGCCGAAGGAGAGGGAGGAGCCATACCAAGGC	120
Db	21	ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly	40
Qy	121	CACAAATACCGTAAGCTTAAAGTTAACCAAGGGGGACCTTTGCCATTTTCTTGGGATATT	180
Db	41	HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle	60
Qy	181	TTGTGCACCACAATTTTCAGTAGGAAGCAAGGTATATGTCAGACGACCCCTGCCGACATACCA	240
Db	61	LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro	80
Qy	241	GACTATAAAAAGCTGTGCATTTTCCTGAAAGGATTTTAAATGGGAAAGGGTTCATGAACTTTGA	300
Db	81	AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu	100
Qy	301	GACGGTGGCTGTTACTGTAACCCACGATTCAGTTTGCAGGATGGCTGTTTCATCTAC	360
Db	101	AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr	120
Qy	361	AAGGTCAAGTTCATTGGCGTGAACCTTTCCTTCCCATGGACCTGTTTATGCAAAAGAAGACA	420
Db	121	LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr	140
Qy	421	ATGGGTGGGAAGCCACGACGTGACGGTTTGTATCCTCGTGATGGCGTGTGTAAGAGGAGAG	480
Db	141	MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu	160
Qy	481	ATTCAATAGGCTCTGAAGCTGAAGACGGTGGTGCATTACCTAGTTGAATTTCAAAGATATT	540
Db	161	IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle	180
Qy	541	TACATGCCAAGAGCCCTGTGCGACTACCGGCTACTACTATGTTGACTCCAAACTGGAT	600
Db	181	TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp	200
Qy	601	ATAACAGCCACACGAAGACTATACAACTCGTTGACGAGTATGAAGAAGACCGAGGAGCCG	660
Db	201	IleThrSerHisAsnGlnAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg	220
Qy	661	CACCATCTGTTTCCTT 675	
Db	221	HisHisLeuPheLeu 225	
RESULT 8			
US-10-315-920-2			
; Sequence 2, Application US/10315920			
; Publication No. US20030175809A1			
; GENERAL INFORMATION:			
; APPLICANT: Pradkov, Arcady Fedorovich			
; APPLICANT: Tersikh, Alexey			
; TITLE OF INVENTION: FLUORESCENT TIMER PROTEINS AND METHODS			
; TITLE OF INVENTION: FOR THEIR USE			
; FILE REFERENCE: CLON-077CIP			
; CURRENT APPLICATION NUMBER: US/10/315,920			
; CURRENT FILING DATE: 2002-12-09			
; PRIOR APPLICATION NUMBER: 60/211,607			
; PRIOR FILING DATE: 2000-06-14			
; PRIOR APPLICATION NUMBER: PCT/US01/19097			
; PRIOR FILING DATE: 2001-06-13			
; NUMBER OF SEQ ID NOS: 22			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 2			
; LENGTH: 225			
; TYPE: PRT			
; ORGANISM: Discosoma sp.			
US-10-315-920-2			

Alignment Scores:

Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 14 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-315-920-2 (1-225)

```
Qy 1 ATGAGGTCTCCAAAGATGTTATCAAGAGTTCATGAGGTTTAAAGGTTTCGATCGAAGGA 60
Db 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAACGCGAAGGAGGAGGCGGACCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTACACAAATTTCAAGTATGGAAGCAAGGTATATGTCAAGACCCCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAGCTGTCAATTTCCCTGAAGGATTTAAATGGGAAGGTCATGAACCTTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGTCTACTGTAAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGTCAAGTTCATGGGTGAACCTTCTCCATGAGACCTGTATGATGCAAAAGATATT 540
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGCGTGGGAAGCCAGCACGACGCTTGTATCTCGATGGCGGTTGCAAGGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAAGGCTCTGAAGCTGAAGAGCGGTGGTCATTTACCTAGTTGAATTCAAAAGTATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGCAAGAAGCCCTGTCAGCTACCGGCTACTATGTTGACTCCAAACTGGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACCAAGCCACAACGAGACTATACAATCCTTCAGCAGTATGAAGAACCGGAGACGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTCTT 675
Db 221 HisHisLeuPheLeu 225
```

RESULT 9

US-10-132-067-4

; Sequence 4, Application US/10132067

; Publication No. US20030203355A1

; GENERAL INFORMATION:

; APPLICANT: Bradbury, Andrew

; APPLICANT: Zeytun, Ahmet

; APPLICANT: Waldo, Geoffrey

; APPLICANT: The Regents of the University of California

; TITLE OF INVENTION: Fluorobodies: Binding Ligands With Intrinsic

; FILE REFERENCE: 021362-000600US

; CURRENT APPLICATION NUMBER: US/10/132,067

; CURRENT FILING DATE: 2002-04-24

; NUMBER OF SEQ ID NOS: 48

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 4

; LENGTH: 225

; TYPE: PRI

; ORGANISM: Discosoma sp.

; FEATURE:

; OTHER INFORMATION: red fluorescent protein (dsRED)

US-10-132-067-4

Alignment Scores:

Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 15 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-132-067-4 (1-225)

```
Qy 1 ATGAGGTCTCCAAAGATGTTATCAAGGAGTTCATGAGGTTTAAAGGTTTCGATCGAAGGA 60
Db 1 MetArgSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAACGCGAAGGAGGAGGCGGACCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACAATACCGTAAAGCTTAAAGTAAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTACACCAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCCCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAGCTGTCAATTTCCCTGAAGGATTTAAATGGGAAGGTCATGAACCTTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGTCTACTGTAAACCCAGGATTTCCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGTCAAGTTCATGGGTGAACCTTCTCCATGAGACCTGTATGATGCAAAAGAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGCGTGGGAAGCCAGCACGACGCTTGTATCTCGATGGCGGTTGCAAGGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAAGGCTCTGAAGCTGAAGAGCGGTGGTCATTTACCTAGTTGAATTCAAAAGTATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGCAAGAAGCCCTGTCAGCTACCGGCTACTATGTTGACTCCAAACTGGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATACCAAGCCACAACGAGACTATACAATCCTTCAGCAGTATGAAGAACCGGAGACGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTCTT 675
Db 221 HisHisLeuPheLeu 225
```

RESULT 10

US-10-370-570-56

; Sequence 56, Application US/10370570

; Publication No. US20030219717A1

; GENERAL INFORMATION:

; APPLICANT: DAHL, Soren Weis et al.

; TITLE OF INVENTION: FLUOROPHORE COMPLEMENTATION PRODUCTS

; FILE REFERENCE: 3759-0130P
; CURRENT APPLICATION NUMBER: US/10/370,570
; CURRENT FILING DATE: 2003-02-24
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 56
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-370-570-56

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 15 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-370-570-56 (1-225)

Qy 1 ATGAGGCTCTCCAAAGATGTTATCAAGGAGTTTCATGAGTTTAAGGTTTCGCATGCAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAAAGCGCAAGAGAGAGGGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACATACCGTAAAGCTTAAGTTAAGCAACCAAGGGGGACCTTGCCTATTTGCTGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTCCGCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAAAGCTGTCTATTTCTGAAGAGTTTAAATGGGAAAGGTCATGAACCTTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGTCTACTGTATTAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTCAAGTTTCATTGGCGTGAATTTCTTCCGATGACCTGTATGCAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTCGTATGGCGTGTGAAAGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAGGCTCTGAAGCTGAAGAGCGGTGGTCATTACCTAGTTGAATTCAAAGATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAGACCTGTGCAGCTACCGGCTACTATGTTGACTCCCAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATAAAGCCCAACAGAGACTATACATCAATCGTTGAGCAGATGAAAGAACCGAGGAGCG 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCTTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 11

US-10-406-618-32

; Sequence 32, Application US/10406618

; Publication No. US20030219814A1
; GENERAL INFORMATION:
; APPLICANT: Wan, David Chi-Cheong
; APPLICANT: Ip, Tsz-Ming
; APPLICANT: The Chinese University of Hong Kong
; TITLE OF INVENTION: No. US20030219814A1 Fluorescent Proteins
; FILE REFERENCE: 016285-34-1
; CURRENT APPLICATION NUMBER: US/10/406,618
; CURRENT FILING DATE: 2003-04-02
; PRIOR APPLICATION NUMBER: US 60/387,968
; PRIOR FILING DATE: 2002-06-11
; PRIOR APPLICATION NUMBER: US 60/370,598
; PRIOR FILING DATE: 2002-04-05
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Spontaneously
; OTHER INFORMATION: Fluorescent protein DaRed.
US-10-406-618-32

Alignment Scores:

Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 15 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-406-618-32 (1-225)

Qy 1 ATGAGGCTCTCCAAAGATGTTATCAAGGAGTTTCATGAGTTTAAGTTCGCATGCAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20
Qy 61 ACGGTCAATGGCGACGAGTTTGAATAGAAAGCGCAAGAGAGGGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Qy 121 CACATACCGTAAAGCTTAAGTTAAGTAAACCAAGGGGGACCTTGCCTATTTGCTGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCAATTTTCAGTATGGAAGCAAGGTATATGTCAAGCACCTTCCGCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAAAGCTGTCTATTTCTGAAGAGTTTAAATGGGAAAGGTCATGAACCTTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGTCTACTGTATTAACCCAGGATTCAGTTTCAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTCAAGTTTCATTGGCGTGAATTTCTTCCGATGACCTGTATGCAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGGCTGGGAAGCCAGCAGCTAGCGTTTGTATCTCTCGTATGGCGTGTGAAAGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAGGCTCTGAAGCTGAAGAGCGGTGGTCATTACCTAGTTGAATTCAAAGATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAGACCTGTGCAGCTACCGGCTACTATGTTGACTCCCAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200

```
QY 601 ATAAACGCCACACGAGACTATACATCGTTGAGCAGTATGAAAGACCGAGGACGC 660
Db 201 ILeThrSerHisAsnGluAsePtyrThrIleValGluGlnTyrgluAthrGluGlyArg 220

QY 661 CACCATCTGTTCCCTT 675
Db 221 HieHisLeuPheLeu 225

RESULT 12
US-10-433-640-13
; Sequence 13, Application US/10433640
; Publication No. US20040115792A1
; GENERAL INFORMATION:
; APPLICANT: Lichtenberg-Frate, Hella
; TITLE OF INVENTION: YEAST STRAIN FOR TESTING THE GENOTOXICITY AND CYTOTOXICITY OF
; FILE REFERENCE: 1487/3
; CURRENT APPLICATION NUMBER: US/10/433,640
; PRIOR FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/14610
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: DE 10061872.3
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-433-640-13

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 16 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-433-640-13 (1-225)

QY 1 ATGAGGCTTCCAAAGAAATGTTATCAAGAGGTTTCATGAGTTTAAGGTTTCGATGGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

QY 61 ACGGTCATATGGCGACGAGTTTGAATAAGAGCGGAAGAGAGGAGCGGCCATACGAAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40

QY 121 CACAATACCGTAAAGCTTAAGGTTAAGGTAACCAAGGGGGACCTTGGCCATTGCTTGGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

QY 181 TTGTCCACCACAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

QY 241 GACTATAAAAAGCTGTCATTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAATTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100

QY 301 GACGGTCGGCGCTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120

QY 361 AAGGTCAAGTTTCATTGGCGTGAACCTTCTCCATGGACCTGTTATGCAAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140

QY 421 ATGGCGTGGGAAGCCAGCAGCTAGCGTTTGTATCTCGTGATGGCGTTGAAAGGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
```

```
QY 481 ATTCATAAGGCTCTGAAGCTGAAGACGGTGGTGCATTACCTAGTTCGAATTCAAAAGTATT 540
Db 161 ILeHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180

QY 541 TACATGGCAAGAGGCTGTGCAGCTACCGAGGTACTACTATATGTTGACTCCAAACTGGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200

QY 601 ATAAACGCCACACGAGACTATACATCGTTGAGCAGTATGAAAGACCGAGGACGC 660
Db 201 ILeThrSerHisAsnGluAsePtyrThrIleValGluGlnTyrgluAthrGluGlyArg 220

QY 661 CACCATCTGTTCCCTT 675
Db 221 HieHisLeuPheLeu 225

RESULT 13
US-10-311-030-7
; Sequence 7, Application US/10311030
; Publication No. US20040171107A1
; GENERAL INFORMATION:
; APPLICANT: Nelson, David
; APPLICANT: Zamiaza, Elize
; APPLICANT: Tsien, Roger
; TITLE OF INVENTION: MODIFIED FLUORESCENT PROTEINS
; FILE REFERENCE: 15916-032US1
; CURRENT APPLICATION NUMBER: US/10/311,030
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: PCT/US01/04625
; PRIOR FILING DATE: 2001-02-13
; PRIOR APPLICATION NUMBER: US 60/184,732
; PRIOR FILING DATE: 2000-02-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: synthetic construct
US-10-311-030-7

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 16 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-311-030-7 (1-225)
```

```
QY 1 ATGAGGCTTCCAAAGAAATGTTATCAAGAGGTTTCATGAGTTTAAGGTTTCGATGGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

QY 61 ACGGTCATATGGCGACGAGTTTGAATAAGAGCGGAAGAGAGGAGCGGCCATACGAGGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40

QY 121 CACAATACCGTAAAGCTTAAGGTTAAGGTAACCAAGGGGGACCTTGGCCATTGCTTGGGATATT 180
Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60

QY 181 TTGTCCACCACAATTTCAAGTATGGAAGCAAGGTATATGTCAAGCACCTGCCGACATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80

QY 241 GACTATAAAAAGCTGTCATTTCTGAAAGGATTTAAATGGGAAAGGGTCATGAATTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
```

QY 301 GACGTCGCGCTTACTGTAACCCAGGATTCAGTTTCAGGATGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGGTCAAGTTCATTCGCGTGAACCTTCCTCCATGAGCTGTATGCAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGGAAGCCAGCAGCTAGCGTTGTATCCTCGTATGCGCTGTGAAAGGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATTCATAGGCTCTGAGCTCAAGAGCGGTGCTCACTTACCTAGTTGTAATTCAAAGATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCAAGAGCCCTGTCAGCTACCAAGGCTACTATGTTGATCTCCAAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATAAAGCCCAACGAGACTATACAACTGTTGAGCAGTATGAAAGAGGAGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 14

US-10-845-484-3
; Sequence 3, Application US/10845484
; Publication No. US20040248180A1
; GENERAL INFORMATION:
; APPLICANT: Bulina, Maria
; APPLICANT: Chudakov, Dmitry
; APPLICANT: Lukyanov, Konstantin
; TITLE OF INVENTION: Mutant Chromophores/Fluorophores and
; FILE OF INVENTION: Methods for Making and Using the Same
; FILE REFERENCE: CLON 092
; CURRENT APPLICATION NUMBER: US/10/845,484
; PRIOR FILING DATE: 2004-05-12
; PRIOR APPLICATION NUMBER: 60/343128
; PRIOR FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: PCT/US02/41418
; PRIOR FILING DATE: 2002-12-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-845-484-3

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 16 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-845-484-3 (1-225)

QY 1 ATGAGGTCTTCCAGAAATGTTATCAAGAGTTCATGAGTTTAAGGTTTCGATCGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleGlyGluPheMetArgPheLysValArgMetGluGly 20
QY 61 ACGTCAATGGCGACGAGTTTGAATAGAGCGCAAGGAGAGGAGCGGAGCCATACGAAGC 120
Db 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyGluGlyArgProTyrGluGly 40
QY 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGAGCTTTGCCATTTGCTTGGGATATT 180

Db 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
QY 181 TTGTCCACCAATTTCAAGTATGAAAGCAAGGTATATGTCAAGCACCTCGCGCATACCA 240
Db 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
QY 241 GACTATATAAAGCTGTCATTTCTCTGAAGGATTTAATATGGGAAAGGCTCATGAACTTTGAA 300
Db 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
QY 301 GACGTCGCGCTGCTTACTGTAAACCCAGGATTCAGTTTCAGGATGCTGTTTCATCTAC 360
Db 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
QY 361 AAGTCAAGTTCATTCGCGTGAACCTTTCTTCCGATGACCTGTTATGCAAAAGAGACA 420
Db 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
QY 421 ATGGCTGGGAAGCCAGCAGCTAGCGTTTCATCTCTGATGCGCTGTGAAAGGAGAG 480
Db 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
QY 481 ATTCATAGGCTCTGAGCTGAAAGACGCGTGGTCATTACCTAGTTGAATTCAAAGATT 540
Db 161 IleHisLysAlaLeuLysLeuLysAspGlyHisTyrLeuValGluPheLysSerIle 180
QY 541 TACATGGCAAGAGCCCTGTCAGCTACCAAGGCTACTATGTTGATCTCCAAACTCGAT 600
Db 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrValAspSerLysLeuAsp 200
QY 601 ATAAAGCCCAACGAGACTATACAACTGTTGAGCAGTATGAAAGAGGAGCGC 660
Db 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
QY 661 CACCATCTGTTCTT 675
Db 221 HisHisLeuPheLeu 225

RESULT 15

US-10-885-988-12
; Sequence 12, Application US/10885988
; Publication No. US20040259165A1
; GENERAL INFORMATION:
; APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: TS'EN, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
; FILE REFERENCE: REGEN1530-2
; CURRENT APPLICATION NUMBER: US/10/885,988
; CURRENT FILING DATE: 2004-07-06
; PRIOR APPLICATION NUMBER: US/09/866,538
; PRIOR FILING DATE: 2001-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Discosoma sp.
US-10-885-988-12

Alignment Scores:
Pred. No.: 2,99e-128 Length: 225
Score: 1214.00 Matches: 225
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 99.92% Indels: 0
DB: 16 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-885-988-12 (1-225)

QY 1 ATGAGGTCTTCCAGAAATGTTATCAAGAGTTCATGAGTTTAAGGTTTCGATCGAAGGA 60
Db 1 MetArgSerSerLysAsnValIleLysGluPheMetArgPheLysValArgMetGluGly 20

```
Qy 61 ACGGTCAATGGGCACGAGTTTGAATAAGCGAAGGAGAGGGAGGCCCATACGAAGGC 120
Db |||||||
Qy 21 ThrValAsnGlyHisGluPheGluIleGluGlyGluGlyArgProTyrGluGly 40
Db |||||||
Qy 121 CACAATACCGTAAAGCTTAAGGTAAACCAAGGGGGACCTTTGCCATTGTCTGGGATATT 180
Db |||||||
Qy 41 HisAsnThrValLysLeuLysValThrLysGlyGlyProLeuProPheAlaTrpAspIle 60
Qy 181 TTGTCCACCAATTTCAAGTATGGAAGCAAGGTATATGTCAGCACCCCTGCCGACATACCA 240
Db |||||||
Qy 61 LeuSerProGlnPheGlnTyrGlySerLysValTyrValLysHisProAlaAspIlePro 80
Qy 241 GACTATAAAGACTGTCTCATTTCTCTGAAGGATTTAAATGGGAAAGGGTCATGAACCTTTGAA 300
Db |||||||
Qy 81 AspTyrLysLysLeuSerPheProGluGlyPheLysTrpGluArgValMetAsnPheGlu 100
Qy 301 GACGGTGGCGCTTACTGTAAACCCAGATTCCAGTTTCAGATGGCTGTTCATCTAC 360
Db |||||||
Qy 101 AspGlyGlyValValThrValThrGlnAspSerSerLeuGlnAspGlyCysPheIleTyr 120
Qy 361 AAGGTCAAGTTTCATTGGCGTGAACCTTCCTCCGATGGACCTGTATGCAAAAGAGACA 420
Db |||||||
Qy 121 LysValLysPheIleGlyValAsnPheProSerAspGlyProValMetGlnLysLysThr 140
Qy 421 ATGGGCTGGGAAGCCAGCACTGAGCGTTTGTATCTCTGATGGCGTGTGAAAGGAGAG 480
Db |||||||
Qy 141 MetGlyTrpGluAlaSerThrGluArgLeuTyrProArgAspGlyValLeuLysGlyGlu 160
Qy 481 ATTCATAAGGCTCTGAAGCTGAAGACGGTGGTCATTACCTAGTTGAATTCAAAAGTATT 540
Db |||||||
Qy 161 IleHisLysAlaLeuLysLeuLysAspGlyGlyHisTyrLeuValGluPheLysSerIle 180
Qy 541 TACATGGCAAGAGCCCTGTCAGCTACCAGGGTACTACTATGTTGACTCCAACTCGAT 600
Db |||||||
Qy 181 TyrMetAlaLysLysProValGlnLeuProGlyTyrTyrTyrValAspSerLysLeuAsp 200
Qy 601 ATAACAAGCCACAAAGCAAGACTATACAAATCGTTGAGCAGTATGAAAGAACCGAGGACGC 660
Db |||||||
Qy 201 IleThrSerHisAsnGluAspTyrThrIleValGluGlnTyrGluArgThrGluGlyArg 220
Qy 661 CACCATCTGTTCCTT 675
Db |||||||
Qy 221 HisHisLeuPheLeu 225
```

Search completed: June 30, 2005, 16:11:22
Job time : 141.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 30, 2005, 15:51:15 ; Search time 32 Seconds
(without alignments)
3163.254 Million cell updates/sec

Title: US-10-006-922A-11

Perfect score: 1215

Sequence: 1 atagggtctccagaagtgt.....gccaccatctgttcctttaa 678

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 1027090

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_pool_P/US10006922/runat 30062005 110553 9450/app query.fasta 1.839
-DB=Issued Patents AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0 1 -LOOPC1=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10006922 @CGN 1 1 46 @runat 30062005 110553 9450 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5A COMB.pep:*
- 2: /cgn2_6/ptodata/1/iaa/5B COMB.pep:*
- 3: /cgn2_6/ptodata/1/iaa/6A COMB.pep:*
- 4: /cgn2_6/ptodata/1/iaa/6B COMB.pep:*
- 5: /cgn2_6/ptodata/1/iaa/6C COMB.pep:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1210	99.6	240	4	US-10-152-296-2
2	485.5	40.0	238	3	Sequence 2, Appli
3	485.5	40.0	238	4	Sequence 16, Appl
4	485.5	40.0	238	4	Sequence 16, Appl
5	485.5	40.0	238	4	Sequence 65, Appl
6	485.5	40.0	238	4	Sequence 65, Appl
7	485.5	40.0	238	4	Sequence 65, Appl
8	485.5	40.0	238	4	Sequence 65, Appl
9	485.5	40.0	238	4	Sequence 65, Appl
10	485.5	40.0	238	4	Sequence 65, Appl
11	485.5	40.0	238	4	Sequence 65, Appl
12	485.5	40.0	238	4	Sequence 65, Appl

13	473.5	39.0	229	4	US-09-977-897-18	Sequence 18, Appl
14	473.5	39.0	230	4	US-09-977-897-19	Sequence 19, Appl
15	473.5	39.0	231	4	US-09-977-897-20	Sequence 20, Appl
16	473.5	39.0	232	4	US-09-977-897-21	Sequence 21, Appl
17	473.5	39.0	233	4	US-09-977-897-22	Sequence 22, Appl
18	473.5	39.0	234	4	US-09-977-897-9	Sequence 9, Appl
19	473.5	39.0	234	4	US-09-977-897-23	Sequence 23, Appl
20	473.5	39.0	235	4	US-09-977-897-8	Sequence 8, Appl
21	473.5	39.0	235	4	US-09-977-897-24	Sequence 24, Appl
22	473.5	39.0	236	4	US-09-977-897-7	Sequence 7, Appl
23	473.5	39.0	236	4	US-09-977-897-25	Sequence 25, Appl
24	473.5	39.0	237	4	US-09-977-897-6	Sequence 6, Appl
25	473.5	39.0	237	4	US-09-977-897-26	Sequence 26, Appl
26	473.5	39.0	237	4	US-09-977-897-27	Sequence 27, Appl
27	473.5	39.0	238	3	US-09-277-716-32	Sequence 32, Appl
28	473.5	39.0	238	4	US-09-609-161B-32	Sequence 32, Appl
29	473.5	39.0	238	4	US-09-839-650-3	Sequence 3, Appl
30	473.5	39.0	238	4	US-09-977-897-5	Sequence 5, Appl
31	473.5	39.0	238	4	US-09-977-897-27	Sequence 27, Appl
32	473.5	39.0	239	4	US-09-977-897-2	Sequence 2, Appl
33	473.5	39.0	239	4	US-09-977-897-4	Sequence 4, Appl
34	473.5	39.0	239	4	US-09-977-897-12	Sequence 12, Appl
35	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
36	473.5	39.0	239	4	US-09-977-897-10	Sequence 10, Appl
37	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
38	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
39	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
40	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
41	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
42	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
43	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
44	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl
45	473.5	39.0	239	4	US-09-977-897-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1

US-10-152-296-2

; Sequence 2, Application US/10152296

; Patent No. 6723537

; GENERAL INFORMATION:

; APPLICANT: Peelle, Beau

; APPLICANT: Rigel Pharmaceuticals, Incorporated

; TITLE OF INVENTION: Directed Evolution of Protein in Mammalian Cells

; FILE REFERENCE: 021044-000110US

; CURRENT APPLICATION NUMBER: US/10152.296

; CURRENT FILING DATE: 2002-12-10

; PRIOR APPLICATION NUMBER: US 60/291,871

; PRIOR FILING DATE: 2001-05-18

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 240

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:mammalian

; OTHER INFORMATION: codon-optimized variant (DSRED) of Discosoma sp.

; OTHER INFORMATION: "red" red fluorescent protein (RFP)

US-10-152-296-2

Alignment Scores:

Pred. No.: 1.16e-137 Length: 240

Score: 1210.00 Matches: 224

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 99.56% Mismatches: 0

Query Match: 99.59% Indels: 0

DB: 4 Gaps: 0

US-10-006-922A-11 (1-678) x US-10-152-296-2 (1-240)

Qy

1 ATGAGGTCTTCCAGAAGTGTATCAAGGAGTTCATGAGTTTAAAGTTCCATCGAAGGA 60

; PRIOR APPLICATION NUMBER: 60/089,367
; PRIOR FILING DATE: 1998-06-15
; PRIOR APPLICATION NUMBER: 60/079,624
; PRIOR FILING DATE: 1998-03-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla mulleri
; FEATURE:
; OTHER INFORMATION: Renilla mulleri Green Fluorescent Protein (GFP)
US-09-609-161B-16

Alignment Scores:
Pred. No.: 7,56e-50 Length: 238
Score: 485.50 Matches: 97
Percent Similarity: 64.22% Conservative: 43
Best Local Similarity: 44.50% Mismatches: 71
Query Match: 39.96% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-609-161B-16 (1-238)

```
Qy 10 TCCAGAAATGTTATCAAG-----GAGTTTCATGAGTTTAAAGGTTTCGCATG 54
Db 2 SerlysglnlleuLysAsnThrCysLeuGlnGluValMetSerTyrLysValAsnLeu 21
Qy 55 GAAGGACCGTCATGCGGACGAGTTTGAATAGAACGCGAAGGAGAGGAGGCCATAC 114
Db 22 GluGlylleValAsnAsnHisValPheThrMetGluGlyCysGlyLysGlyAsnIleLeu 41
Qy 115 GAAGGCCCAATACCGTTAAAGTAAACCAAGCGGAGCCCTTTGCCATTTGCTGG 174
Db 42 PheGlyAsnGlnLeuValGlnIleArgValThrLysGlyAlaProLeuProPheAlaPhe 61
Qy 175 GATATTTTGCACCAATTTCAATGTTGAAGCAAGATATATGTCAGACACCTGCGGAC 234
Db 62 AspIleValSerProAlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAsnAsp 81
Qy 235 ATACCAGACTATAAAGCTGTCTATTCCTCAAGGATTTAAATGGGAAGGTCATGAAC 294
Db 82 IleSerAspTyrPheIleGlnSerPheProAlaGlyPheMetTyrGluArgThrLeuArg 101
Qy 295 TTTGAAGACGGTGGCGTCTACTGTAAACCCAGGATTCAGTTTCCAGGATGGCTGTTTC 354
Db 102 TyrGluaspGlyGlyLeuValGluIleArgSerAspIleAsnLeuIleGluAspLysPhe 121
Qy 355 ATCTCAAGGTCAGTTTCATTTGCGTGAACCTTCCCTCCGATGGACCTGTTATGCAAA 414
Db 142 ThrIleLeuGlyIleGluProSerPheGluAlaMetTyrMetAsnAsnGlyValLeuVal 161
Qy 475 GGAGAGATTCATAAGGCTGTAAGCTGAACGCGGATTCATTAAGTTGAATTCAAA 534
Db 162 GlyGluValIleLeuValTyrLysLeuAsnSerGlyLysTyrTyrSerCysHisMetLys 181
Qy 535 AGTATTTACATGGCAAGACCTGTG---CAGCTACAGGTTACTACTATGTTGACTCC 591
Db 182 ThrLeuMetLysSerLysGlyValValLysGluPheProSerTyrHisPheIleGlnHis 201
Qy 592 AAATCGATATAACAGCCCAAGACATATACAAATCGTTGAGCAGTATGAA 645
Db 202 ArgLeuGluLysThr---TyrValGluAspGlyGlyPheValGluGlnHisGlu 218
```

RESULT 4

US-09-626-581D-65
; Sequence 65, Application US/09626581D
; Patent No. 6548249
; GENERAL INFORMATION:

; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; FILE REFERENCE: A-66900-3/RMS
; CURRENT APPLICATION NUMBER: US/09/626,581D
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; PRIOR APPLICATION NUMBER: 09/415,765
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-581D-65

Alignment Scores:
Pred. No.: 7,56e-50 Length: 238
Score: 485.50 Matches: 97
Percent Similarity: 64.22% Conservative: 43
Best Local Similarity: 44.50% Mismatches: 71
Query Match: 39.96% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-626-581D-65 (1-238)

```
Qy 10 TCCAGAAATGTTATCAAG-----GAGTTTCATGAGTTTAAAGGTTTCGCATG 54
Db 2 SerlysglnlleuLysAsnThrCysLeuGlnGluValMetSerTyrLysValAsnLeu 21
Qy 55 GAAGGACCGTCATGCGGACGAGTTTGAATAGAACGCGAAGGAGAGGAGGCCATAC 114
Db 22 GluGlylleValAsnAsnHisValPheThrMetGluGlyCysGlyLysGlyAsnIleLeu 41
Qy 115 GAAGGCCCAATACCGTTAAAGTAAACCAAGCGGAGCCCTTTGCCATTTGCTGG 174
Db 42 PheGlyAsnGlnLeuValGlnIleArgValThrLysGlyAlaProLeuProPheAlaPhe 61
Qy 175 GATATTTTGCACCAATTTCAATGTTGAAGCAAGATATATGTCAGACACCTGCGGAC 234
Db 62 AspIleValSerProAlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAsnAsp 81
Qy 235 ATACCAGACTATAAAGCTGTCTATTCCTCAAGGATTTAAATGGGAAGGTCATGAAC 294
Db 82 IleSerAspTyrPheIleGlnSerPheProAlaGlyPheMetTyrGluArgThrLeuArg 101
Qy 295 TTTGAAGACGGTGGCGTCTACTGTAAACCCAGGATTCAGTTTCCAGGATGGCTGTTTC 354
Db 102 TyrGluaspGlyGlyLeuValGluIleArgSerAspIleAsnLeuIleGluAspLysPhe 121
Qy 355 ATCTCAAGGTCAGTTTCATTTGCGTGAACCTTCCCTCCGATGGACCTGTTATGCAAA 414
Db 122 ValTyrArgValGluTyrLysGlySerAsnPheProAspGlyProValMetGlnLys 141
Qy 415 AAGACATGGCTGGGAGCGGACCTGAGCTGTTGATCTCTCGTGGCGTGTGAAA 474
Db 142 ThrIleLeuGlyIleGluProSerPheGluAlaMetTyrMetAsnAsnGlyValLeuVal 161
Qy 475 GGAGAGATTCATAAGGCTGTAAGCTGAACGCGGATTCATTAAGTTGAATTCAAA 534
Db 162 GlyGluValIleLeuValTyrLysLeuAsnSerGlyLysTyrTyrSerCysHisMetLys 181
Qy 535 AGTATTTACATGGCAAGACCTGTG---CAGCTACAGGTTACTACTATGTTGACTCC 591
Db 182 ThrLeuMetLysSerLysGlyValValLysGluPheProSerTyrHisPheIleGlnHis 201
Qy 592 AAATCGATATAACAGCCCAAGACATATACAAATCGTTGAGCAGTATGAA 645
Db 202 ArgLeuGluLysThr---TyrValGluAspGlyGlyPheValGluGlnHisGlu 218
```

RESULT 5

```
US-09-415-765B-65
; Sequence 65, Application US/09415765B
; Patent No. 6548632
; GENERAL INFORMATION:
; APPLICANT: Anderson, David
; TITLE OF INVENTION: Fusions of Scaffold Proteins with Random Peptide
; FILE REFERENCE: A66900-1/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/415,765B
; CURRENT FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-415-765B-65

Alignment Scores:
Pred. No.: 7,56e-50 Length: 238
Score: 485.50 Matches: 97
Percent Similarity: 64.22% Conservative: 43
Best Local Similarity: 44.50% Mismatches: 71
Query Match: 39.96% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-415-765B-65 (1-238)
Qy 10 TCCAGAAATGTTATCAAG-----GAGTTTCATGAGTTTAAAGTTTCCGATG 54
Db 2 SerLysGlnIleLeuLysAsnThrCysLeuGlnGluValMetSerTyrLysValAsnLeu 21
Qy 55 GAAGGACCGTCAATGGCAGCAGTTTCAATAGACGAGGAGGAGGAGGAGGCCATAC 114
Db 22 GluGlyIleValAsnAsnHisValPheThrMetGluGlyCysGlyLysGlyAsnIleLeu 41
Qy 115 GAAGGCCACAATACCGTAAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTGCTTGG 174
Db 42 PheGlyAsnGlnLeuValGlnIleArgValThrLysGlyAlaProLeuProPheAlaPhe 61
Qy 175 GATATTTTGCACCAATTTTCAGTATGGAAGCAAGGTATATGTCACGACCTTCCGAC 234
Db 62 AspIleValSerProAlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAsnAsp 81
Qy 235 ATACCAAGCTATAAAGCTGCTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTC 354
Db 102 TyrGluaspGlyLeuValGluIleArgSerAspIleAsnLeuIleGluaspLysPhe 121
Qy 355 ATCTCAAGGTCAAGTTTCATTTCCGATGACCTTTCCGATGACCTGTTATGCAAAAG 414
Db 122 ValTyrArgValGluTyrLysGlySerAsnPheProaspGlyProValMetGlnLys 141
Qy 415 AAGACATGGCTGGGAGCCGACCTGAGCGTTTGTATCTCTGATGCGGTGTTGAAA 474
Db 142 ThrIleLeuGlyIleGluProSerPheGluAlaMetTyrMetAsnAsnGlyValLeuVal 161
Qy 475 GGAGAGATTCAATAGGCTGTAAGCTGAAGACCGTGGTTCATTACCTAGTTGAATTCAAA 534
Db 162 GlyGluValIleLeuValTyrLysLeuAsnSerGlyLysTyrTyrSerCysHisMetLys 181
Qy 535 AGTATTTACATGCAAGACCTGTG---CAGTTACAGGGGTACTACTATGTTGACTCC 591
Db 182 ThrLeuMetLysSerLysGlyValValLysGluPheProSerTyrHisPheIleGlnHis 201
Qy 592 AAACGTGATATAACGACCAAGACATATACATCTGTTGAGCAGTATGAA 645
Db 202 ArgLeuGlnLysThr---TyrValGluaspGlyGlyPheValGluGlnHisGlu 218

US-09-626-580C-65
; Sequence 65, Application US/09626580C
; Patent No. 6562617
; GENERAL INFORMATION:
; APPLICANT: Bogenberger, David
; TITLE OF INVENTION: FUSIONS OF SCAFFOLD PROTEINS WITH RANDOM PEPTIDE LIBRARIES
; FILE REFERENCE: A-66900-2/RMS/AMS
; CURRENT APPLICATION NUMBER: US/09/626,580C
; CURRENT FILING DATE: 2000-07-27
; PRIOR APPLICATION NUMBER: US 09/415,765
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: US 09/169,015
; PRIOR FILING DATE: 1998-10-08
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 65
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Renilla muelleri
US-09-626-580C-65

Alignment Scores:
Pred. No.: 7,56e-50 Length: 238
Score: 485.50 Matches: 97
Percent Similarity: 64.22% Conservative: 43
Best Local Similarity: 44.50% Mismatches: 71
Query Match: 39.96% Indels: 7
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-626-580C-65 (1-238)
Qy 10 TCCAGAAATGTTATCAAG-----GAGTTTCATGAGTTTAAAGTTTCCGATG 54
Db 2 SerLysGlnIleLeuLysAsnThrCysLeuGlnGluValMetSerTyrLysValAsnLeu 21
Qy 55 GAAGGACCGTCAATGGCAGCAGTTTGAATAGACGAGGAGGAGGAGGAGGCCATAC 114
Db 22 GluGlyIleValAsnAsnHisValPheThrMetGluGlyCysGlyLysGlyAsnIleLeu 41
Qy 115 GAAGGCCACAATACCGTAAAGCTTAAGTAAACCAAGGGGGACCTTTGCCATTGCTTGG 174
Db 42 PheGlyAsnGlnLeuValGlnIleArgValThrLysGlyAlaProLeuProPheAlaPhe 61
Qy 175 GATATTTTGCACCAATTTTCAGTATGGAAGCAAGGTATATGTCACGACCTTCCGAC 234
Db 62 AspIleValSerProAlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAsnAsp 81
Qy 235 ATACCAAGCTATAAAGCTGCTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTC 294
Db 82 IleSerAspTyrPheIleGlnSerPheProAlaGlyPheMetTyrGluArgThrLeuArg 101
Qy 295 TTTGAAGACGGTGGCTGCTGCTTACTGTAACCCAGGATTCAGTTTGCAGGATGGCTGTTTC 354
Db 102 TyrGluaspGlyLeuValGluIleArgSerAspIleAsnLeuIleGluaspLysPhe 121
Qy 355 ATCTCAAGGTCAAGTTTCATTTCCGATGACCTTTCCGATGACCTGTTATGCAAAAG 414
Db 122 ValTyrArgValGluTyrLysGlySerAsnPheProaspGlyProValMetGlnLys 141
Qy 415 AAGACATGGCTGGGAGCCGACCTGAGCGTTTGTATCTCTGATGCGGTGTTGAAA 474
Db 142 ThrIleLeuGlyIleGluProSerPheGluAlaMetTyrMetAsnAsnGlyValLeuVal 161
Qy 475 GGAGAGATTCAATAGGCTGTAAGCTGAAGACCGTGGTTCATTACCTAGTTGAATTCAAA 534
Db 162 GlyGluValIleLeuValTyrLysLeuAsnSerGlyLysTyrTyrSerCysHisMetLys 181
Qy 535 AGTATTTACATGCAAGACCTGTG---CAGTTACAGGGGTACTACTATGTTGACTCC 591
Db 182 ThrLeuMetLysSerLysGlyValValLysGluPheProSerTyrHisPheIleGlnHis 201
Qy 592 AAACGTGATATAACGACCAAGACATATACATCTGTTGAGCAGTATGAA 645
Db 202 ArgLeuGlnLysThr---TyrValGluaspGlyGlyPheValGluGlnHisGlu 218
```


			:::		:	:::	
Db	LysGlyGlyValGluPheProGluTyrHisPheIleHisArgLeuGluLysThr	645					
QY	ACCCACAGCAACTATCATCGTTGCAGCAGTATCAA	645					
	: :::	:::					
Db	--TyrValGluGluGlySerPheValGluGlnHisGlu	219					

RESULT 8

```

RESULT 8
US-09-977-897-13
; Sequence 13, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green
; TITLE OF INVENTION: protein with codon preference of mammalian expression
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-13

```

Alignment Scores:	
Pred. No.:	2.08e-48
Score:	473.50
Percent Similarity:	64.32%
Best Local Similarity:	45.54%
Query Match:	38.97%
DB:	4
	3
	3
Length:	224
Matches:	97
Conservative:	40
Mismatches:	73
Indels:	3
Gaps:	3

US-10-006-922A-11 (1-678) x US-09-977-897-13 (1-224)

QY 13 AAGATGTT---ATCAAGGAGTTCATGAGTTTAAGGTTCCGATCGAAGAACGGTCAAT 69
:::|||||
Db 8 LysAsnThrGlyLeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsn 27
:::|||||
QY 70 GGCACGAGTTTGAAATAGAACGCCGAAGGAGGCGGAGGCCATACGAAGSCCAACAATACC 129
:::|||||
Db 28 AsnHisValPheSerMetGluGlyPheGlyGlyAsnValLeuPheGlyAsnGlnLeu 47
:::|||||
QY 130 GTAAAGCTTAAGGTAAACCAAGGGGGACCTTGCATTTCTCTGGGATATTTTGTCCACCA 189
:::|||||
Db 48 MetGlnIleArgValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIle 67
:::|||||
QY 190 CAATTTCAGTAGAAGCAAGGTATATGTCAGCACCTCCGACATACACAGACTATAA 249
:::|||||
Db 68 AlaPheGlnTyrgLyAsnArgThrPheThyLysTyProAspAspIleAlaAspTyrPhe 87
:::|||||
QY 250 AAGCTGTCATTTCTCGTAGGAGTTTAAATGGCAAAGGTCATGAACCTTTGAAGACGCTGGC 309
:::|||||
Db 88 ValGlnSerPheProAlaGlyPhePheTyrgLuarAsnLeuArgPheGluAspGlyAla 107
:::|||||
QY 310 CTCGTTACTGTAAACCCAGGATTCAGTTTCAGAGATGGCTGTTTCATCTACAAGGTCAAG 369
:::|||||
Db 108 IleValAspIleArgSerAspIleSerLeuGluAspAspLysPheHisTyryLysValGlu 127
:::|||||
QY 370 TTCATTGGCGTGAACCTTTCCTCCGATGGACCTCTTATGCAAAAGAAGACAATGGGCTGG 429
:::|||||
Db 128 TyrArgGlyAsnGlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMet 147
:::|||||
QY 430 GAAGCCAGCACTGAGCGGTTTATCTCCTCGTGGCGGTGTTGAAGAGGAGAGATTCATAAG 489
:::|||||
Db 148 GluProSerPheGluValValTyMetAsnSerGlyValLeuValGlyGluValAspLeu 167
:::|||||
QY 490 GCTCTGAAGCTGAAGACGGTGGTCATTACTGTTGAAATTCAAAAGTATTTACATGGCA 549
:::|||||
Db 168 ValTyryLysLeuGluSerGlyAsnTyrySer-CysHisMetLysThrPheTyrgArgSer 187
:::|||||

```
QY 550 AAGAAGCCTGTG---CAGTACCAGGGTACTACTACTGTGTGACTCCAACTGGATATAACA 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 LysGlyGlyVallysGluPheProGluTyrHisPheIleHisHsArgLeuGluLysThr 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 607 AGCCACAACGAGACTATACAAATCGTTGAGCAGTATGAA 645
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 ---TyrValGluGluGlySerPheValGluGlnHisGlu 219

RESULT 9
US-09-977-897-14
; Sequence 14, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-14

Alignment Scores:
Pred. No.: 2,09e-48 Length: 225
Score: 473.50 Matches: 97
Percent Similarity: 64.32% Conservative: 40
Best Local Similarity: 45.54% Mismatches: 73
Query Match: 38.97% Indels: 3
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-977-897-14 (1-225)
QY 13 AAGAAATGTT---ATCAAGGAGTTCATGAGGTTTAAAGTTCGCATGGAAGGACGGTCAAT 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 LysAsnThrGlyLeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsn 27
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 GGCACAGGTTTGAATAGAGGCGAAGGAGGAGGAGGCCATACGAGGCGCACCAATACC 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 AsnHisValPheSerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeu 47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 GTAAAGCTTAAGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTACCA 189
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 MetGlnIleArgValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIle 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 CAATTTTCAGTATGCAAGCAGGTATATCTCAAGCACCTCGCGACATACCAGACTATAAA 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 AlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAspAspIleAlaAspTyrPhe 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 AAGCTGTCATTTCCTCAAGGATTTAAATGGGAAAGGTCATGAATTTTGAAGACGGTGGC 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 ValGlnSerPheProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAla 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 GTCGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTCAAGAGTCAAG 369
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 IleValAspIleArgSerAspIleSerLeuGluAspAspLysPheHisTyrLysValGlu 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 TTCATTGGCGTGAACCTTCCTCCGATGGACCTCTATGCAAAAGACACAAATGGGCTGG 429
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 TyrArgGlyAsnGlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMet 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 GAAGCCAGCACTGAGCGTTTGTATCTCGTGATGGCGTGTGTAAGAGGAGAGATTCATAAG 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 GluProSerPheGluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeu 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 GCTCTGAAGCTGAAGACGGTGGTCATTACCTAGTTGAATTCAAAGATTTTACATGGCA 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 ValTyrLysLeuGluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSer 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 550 AAGAAGCCTGTG---CAGTACCAGGGTACTACTACTGTGTGACTCCAACTGGATATAACA 606
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 LysGlyGlyVallysGluPheProGluTyrHisPheIleHisHsArgLeuGluLysThr 207
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 607 AGCCACAACGAGACTATACAAATCGTTGAGCAGTATGAA 645
    :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 208 ---TyrValGluGluGlySerPheValGluGlnHisGlu 219

RESULT 10
US-09-977-897-15
; Sequence 15, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluoresce
; TITLE OF INVENTION: protein with codon preference of mammalian expression systems an
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 226
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-15

Alignment Scores:
Pred. No.: 2,09e-48 Length: 226
Score: 473.50 Matches: 97
Percent Similarity: 64.32% Conservative: 40
Best Local Similarity: 45.54% Mismatches: 73
Query Match: 38.97% Indels: 3
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-977-897-15 (1-226)
QY 13 AAGAAATGTT---ATCAAGGAGTTCATGAGGTTTAAAGTTCGCATGGAAGGACGGTCAAT 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8 LysAsnThrGlyLeuLysGluIleMetSerAlaLysAlaSerValGluGlyIleValAsn 27
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 GGCACAGGTTTGAATAGAGGCGAAGGAGGAGGAGGCCATACGAGGCGCACCAATACC 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 28 AsnHisValPheSerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeu 47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 GTAAAGCTTAAGTAACCAAGGGGGACCTTTGCCATTTGCTTGGGATATTTTGTACCA 189
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 48 MetGlnIleArgValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIle 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 CAATTTTCAGTATGCAAGCAGGTATATCTCAAGCACCTCGCGACATACCAGACTATAAA 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 68 AlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAspAspIleAlaAspTyrPhe 87
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 AAGCTGTCATTTCCTCAAGGATTTAAATGGGAAAGGTCATGAATTTTGAAGACGGTGGC 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 88 ValGlnSerPheProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAla 107
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 310 GTCGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTCAAGAGTCAAG 369
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 108 IleValAspIleArgSerAspIleSerLeuGluAspAspLysPheHisTyrLysValGlu 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 370 TTCATTGGCGTGAACCTTCCTCCGATGGACCTCTATGCAAAAGACACAAATGGGCTGG 429
    ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 128 TyrArgGlyAsnGlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMet 147
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 430 GAAGCCAGCACTGAGCGTTTGTATCTCGTGATGGCGTGTGTAAGAGGAGAGATTCATAAG 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 148 GluProSerPheGluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeu 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 490 GCTCTGAAGCTGAAGACGGTGGTCATTACCTAGTTGAATTCAAAGATTTTACATGGCA 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 168 ValTyrLysLeuGluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSer 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


Db 148 GluProSerPheGluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeu 167
Qy 490 GCTCTGAAGCTGAAGACGGTGGTCATTACCTAGTTGAATTCAAAGTATTATTACATGGCA 549
Db 168 ValTyrLysLeuGluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSer 187
Qy 550 AAGAAGCCTGTG---CAGCTACCAAGGCTACTACTATGTTGACTCCAACTGGATATAACA 606
Db 188 LysGlyGlyValLysGluPheProGluTyrHisPheIleHisArgLeuGluLysThr 207
Qy 607 AGCCACAACGAAGACTATACATCGTTGAGCAGTATGAA 645
Db 208 ---TyrValGluGluGlySerPheValGluGlnHisGlu 219

RESULT 15
US-09-977-897-20
; Sequence 20, Application US/09977897
; Patent No. 6780974
; GENERAL INFORMATION:
; APPLICANT: Chen, Yih-Tai
; APPLICANT: Cao, Longguang
; TITLE OF INVENTION: A synthetic DNA encoding an orange seapen-derived green fluorescent protein with codon preference of mammalian expression systems and
; FILE OF INVENTION: protein with codon preference of mammalian expression systems and
; FILE REFERENCE: 41856-5
; CURRENT APPLICATION NUMBER: US/09/977,897
; CURRENT FILING DATE: 2001-10-15
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Ptilosarcus gurneyi
US-09-977-897-20

Alignment Scores:
Pred. No.: 2,12e-48 Length: 231
Score: 473.50 Matches: 97
Percent Similarity: 64.32% Conservative: 40
Best Local Similarity: 45.54% Mismatches: 73
Query Match: 38.97% Indels: 3
DB: 4 Gaps: 3

US-10-006-922A-11 (1-678) x US-09-977-897-20 (1-231)

Qy 13 AAGATGTT---ATCAAGAGTTCATGAGGTTTAAAGTTTCGATCGAAGGAACGGTCAAT 69
Db 8 LysAsnThrGlyLeuLysGluLeuMetSerAlaLysAlaSerValGluGlyIleValAsn 27
Qy 70 GGGCAGGAGTTTGAATAAGAGCGGAGAGGAGGGAGGCCATACGAGGCCCAATACC 129
Db 28 AsnHisValPheSerMetGluGlyPheGlyLysGlyAsnValLeuPheGlyAsnGlnLeu 47
Qy 130 GTAAGCTTAAGGTAACCAAGGGGACCTTTGCTTCATTTGGATATTTCTCACC 189
Db 48 MetGlnIleArgValThrLysGlyGlyProLeuProPheAlaPheAspIleValSerIle 67
Qy 190 CAATTTCAGTATGAAGCAAGCATATCTCAAGCACCTCGCCGACATACCACTATATAA 249
Db 68 AlaPheGlnTyrGlyAsnArgThrPheThrLysTyrProAspAspIleAlaAspTyrPhe 87
Qy 250 AAGCTGTCATTTCTGAAGGATTTAAATGGAAAGGGTCATGAACCTTTGAAGACGGTGGC 309
Db 88 ValGlnSerPheProAlaGlyPhePheTyrGluArgAsnLeuArgPheGluAspGlyAla 107
Qy 310 GTCGTTACTGTAAACCCAGGATTCAGTTTGCAGGATGGCTGTTTCATCTACAGGTCAG 369
Db 108 IleValAspIleArgSerAspIleSerLeuGluAspLysPheHisTyrLysValGlu 127
Qy 370 TTCATTGCGGTGAACCTTTCTTCGACCTGTTATGCAAAAGAACACATGGGTGG 429
Db 128 TyrArgGlyAsnGlyPheProSerAsnGlyProValMetGlnLysAlaIleLeuGlyMet 147
Qy 430 GAAGCCAGCACTGAGCGGTTTGTATCTCGTGTATCGCGGTGTTGAAGGAGGATTCATAAG 489

Db 148 GluProSerPheGluValValTyrMetAsnSerGlyValLeuValGlyGluValAspLeu 167
Qy 490 GCTCTGAAGCTGAAGACGGTGGTTCATTACCTAGTTGAATTCAAAGTATTATTACATGGCA 549
Db 168 ValTyrLysLeuGluSerGlyAsnTyrTyrSerCysHisMetLysThrPheTyrArgSer 187
Qy 550 AAGAAGCCTGTG---CAGCTACCAAGGCTACTACTATGTTGACTCCAACTGGATATAACA 606
Db 188 LysGlyGlyValLysGluPheProGluTyrHisPheIleHisArgLeuGluLysThr 207
Qy 607 AGCCACAACGAAGACTATACATCGTTGAGCAGTATGAA 645
Db 208 ---TyrValGluGluGlySerPheValGluGlnHisGlu 219

Search completed: June 30, 2005, 16:06:40
Job time : 36 secs

THIS PAGE BLANK (USP10)